

Stem-cell research and the US Congress

US Senator Arlen Specter is providing much-needed leadership in Washington on the volatile issue of stem-cell research, as the Clinton administration keeps a safe distance.

The US Senate is considering the circumstances under which federal funds can be used for research on embryonic stem cells, or to extract such cells from human embryos. During the debate, opponents of federal funding for such research have suggested that researchers turn their attention instead to adult stem cells (see page 6). But the adult-derived cells cannot yet be directed into as many different cell types as can embryonic stem cells, and are less inclined to replicate. As Senator Arlen Specter (Republican, Pennsylvania) has pointed out, it would be short-sighted for the Congress to close the door on publicly funded embryonic stem-cell research before it can be more fully explored.

The prospect of such research has set off alarm bells in the powerful anti-abortion movement in the United States. Specter's track record in opposing abortion therefore strengthens his hand in proposing legislation that would authorize the federal government to fund both the use and derivation of embryonic stem cells.

Last week, during hearings of the appropriations subcommittee that he chairs, Specter took testimony from opponents of such funding, but quickly rejected several of their arguments. Embryos left over from *in vitro* fertilization attempts should not be regarded as potential human beings, he argued, as the embryos would only be discarded — as is current practice — if they were not used in research. The embryonic stem cells appear to be more versatile than their adult counterparts, Specter added. He also chastised Senator Sam Brownback (Republican, Kansas) for attempting to equate stem-cell extraction and research with Nazi experiments.

Specter's sponsorship of the legislation means that it has a chance of passing, at least in the Senate. Its progress to date reflects his strong feelings on the issue, as well as some political dexterity. He had initially attached the legislation to an appropriations bill last year. Since its inclusion was likely to bog down the budget process, Trent Lott (Republican, Mississippi), the Senate majority leader, reached a deal: if Specter removed the stem-cell provision from the budget bill, Lott promised to allow his stem-cell bill onto the Senate floor.

The bill faces an uncertain fate there later this month. But other anti-abortionists in the Senate have indicated they will support it. Even if the bill stalls in the Senate — or, as is more likely, gets lost in the House of Representatives — its progress could prove helpful. In June, the National Institutes of Health plans to implement guidelines to allow experimentation with stem cells, but not their extraction, using public funds. A Senate victory for the Specter bill will help to protect these guidelines from an inevitable challenge in the Congress.

The debate has also illustrated the pressing need for a federal policy on stem cells. Both supporters and opponents of stem-cell research have pointed out that such research is being done, without regulation, in the private sector. The administration has been largely silent on this issue since last May, when the National Bioethics Advisory Commission recommended that the government fund both embryonic stem-cell extraction and research. This silence reflected badly on the influence of the commission, and was seen in some quarters as craven. But as Specter attempts to steer the issue through the Congress, it may prove to have been judicious. ■

Agricultural research whistling in the dark

Publicly funded agricultural research continues to languish in the United States, despite attempts for its stronger support.

An evaluation published this week by the National Academy of Sciences (see page 9) is only the latest in a long line of reports highlighting the inadequate level of support for properly reviewed, basic agricultural research in the United States. The academy says that funding for the National Research Initiative at the US Department of Agriculture (USDA), which supports competitively reviewed university research grants, should be expanded from this year's level of \$119 million to \$500 million.

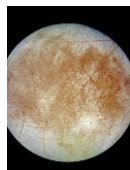
The programme should be elevated in status within the department, the academy panel suggests. Grant awards should be increased to a realistic level and a 19 per cent ceiling on research overhead costs — which serves only to deter some researchers from applying for grants at all — should be removed. An outside committee of advisers should guide the programme, and USDA should appoint a permanent chief scientist to replace the current part-time arrangement.

All of this may be laudable. But it is doubtful if those in control of the purse-strings at USDA are paying much attention. Domestic and

international programmes in agricultural research are being widely neglected, despite their rich scientific potential. In the United States, as in most industrialized countries, the government's budget for agriculture is under pressure. Repeated efforts to bolster investment in research have fallen victim to this pressure. The result has been little public investment in plant genomics, for example, to the detriment of both farmers and consumers.

The only way forward is to strengthen the scientific activities at USDA itself. This is recognized by the administration and the Senate, but not by the agriculture appropriations subcommittee of the House of Representatives. Last year, the Clinton administration asked for \$200 million for USDA's National Research Initiative and obtained \$119 million, not the \$500 million envisaged when the initiative was launched in 1991. This year the administration is asking for only \$150 million, indicating that its ambitions for the initiative are on hold. Next year brings a new administration, a new Congress and, hopefully, a more constructive approach to agricultural research. ■





Missions delayed
Technical problems
are disrupting
NASA's schedules
p4



Biotech debate
New European panel
aims to revive public
faith in science
p5



Lines of attack
Differences aired in
stem-cell research
controversy
p6



Prion negative
Survey fails to find
CJD protein in
tissue samples
p7

NIH opposes plans for patenting 'similar' gene sequences

Washington

Sharp differences have emerged between the US National Institutes of Health (NIH) and the Patent and Trademark Office (USPTO) over gene patenting. The disagreement centres on whether a patent should be granted on a genomic sequence of unknown function merely because it is similar to a separate sequence whose function is already understood.

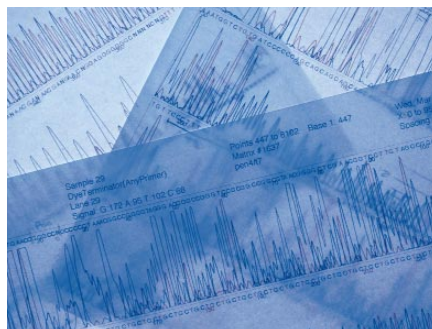
USPTO officials argue that patents should be granted on such a 'homologous' sequence if the two sequences are sufficiently similar to make it likely that the biological function of the product of the new sequence can be predicted with a high degree of confidence. But NIH officials, backed by groups such as the Association of American Medical Colleges (AAMC), are challenging this interpretation.

Although such logic may be applicable to products such as chemical compounds — where two compounds with a similar structure can reasonably be expected to have similar properties — NIH officials argue that a difference in a single base pair in a gene sequence can have important functional implications.

The USPTO is currently revising its rules for awarding gene patents to apply stricter criteria in judging whether an invention can be considered genuinely useful (see *Nature* 403, 3; 2000). Most of the changes, released in draft form just before Christmas, have been widely welcomed. They would, for example, mean that it was no longer possible to describe the utility of an expressed sequence tag as being merely to 'fish' for genes of unknown function.

But the proposals on sequence homology have drawn criticism. In its comments on the draft, for example, the AAMC argues that researchers can often use automated programs to 'guess' the identity and function of a protein encoded by a gene based on the similarity of a fragment to other known genes. "Such suppositions of utility are technology driven, and require little scientific insight or creativity," writes AAMC president Jordan J. Cohen.

The biotech industry appears divided on the homology proposal. While some



Sequence suspicions: questions remain over when genomic data should be patentable.

genomics companies would continue to benefit from the granting of patents based primarily on homology — as was the case recently, for example, when Human Genome Sciences was issued a patent on the CCR5 chemokine receptor later found to be the key receptor for HIV (see *Nature* 404, 322; 2000) — others apparently feel that it could stifle innovation in the field.

But the NIH is unequivocal in its comments. Jack Spiegel, director of the NIH's division of technology transfer and develop-

ment, argues that it is extremely difficult to make an accurate prediction of the biological function of a protein solely on the basis of the similarity of its sequence to another one.

"Minor changes in the nucleotide or amino acid sequences of [...] molecules may produce profound changes in biological activity," writes Spiegel, adding that "homology in an unpredictable art cannot, by itself, provide a specific utility".

The National Advisory Council for Human Genome Research puts it even more bluntly: "Finding partial sequence similarity is an obvious and non-inventive step." Other critics point out that, if patents are granted based solely on homology, the patent holders will have little incentive to continue to a full characterization of the gene product — but could claim the rights to the results of other researchers who later did this.

It remains to be seen how far the USPTO takes on board such criticisms. Although the new guidelines are already being implemented in judging patent applications, a final version will be published within a few months. The USPTO says it is not anticipating "major changes" from the current draft. **David Dickson**

Shareholder sues Celera over loss

Washington

PE Corporation, the parent company of Celera Genomics, is facing legal action from one of Celera's shareholders. The investor claims that information published by Celera misled him into losing money during the recent collapse in the company's share price. Both companies are dismissing the charge as baseless.

The class action lawsuit alleges that when PE Corporation was seeking investors in a 'secondary offering' at the end of February, it failed to disclose the existence of negotiations between Celera and principals in the public Human Genome Project (HGP).

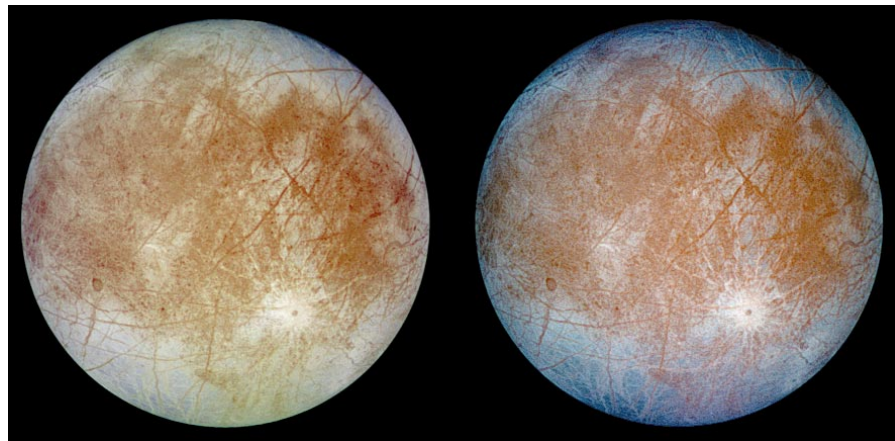
If the case advances, it will investigate a tumultuous time in Celera's stock prices. On 29 February, the company sold 4.37 million

shares at \$225 each. For a week, the share price rose, peaking at nearly \$260 a share.

But negotiations over a possible collaboration between Celera and the HGP on sequencing the human genome broke off about a week later. The impasse — and rancour behind it — soon became public and Celera's stock began slipping. The share price then plummeted on 14 March, triggered by media reports of a joint statement by US President Bill Clinton and British Prime Minister Tony Blair backing the HGP's free information policy (see *Nature* 404, 324; 2000).

PE Corporation says it has done nothing wrong. "We feel that the suit is totally without merit," says Lyn Christianson, a spokesperson for the company. **Paul Smaglik**

Doubts and uncertainties slow NASA's schedule



A late launch: the next mission to Europa (shown in natural, left, and false colour) faces delays.

Washington

In NASA's struggle to balance 'better, faster and cheaper', faster is starting to lose out. The space agency is considering delaying the launch dates for a number of high-profile missions on its wish list — including an orbiter to explore Jupiter's icy moon Europa.

The Jet Propulsion Laboratory (JPL) in Pasadena, which is planning the Europa project, has recommended that the launch date be moved from November 2003 to no earlier than January 2006. This rescheduling has been forced by uncertainties about the spacecraft's launch vehicle and some of the new technologies it will require.

Placing a spacecraft in orbit around one of Jupiter's moons is more difficult than orbiting the planet, and requires a heavier version of today's Delta or Atlas rockets. The US military has contracted industry to build these so-called evolved expendable launch vehicles. But development is behind schedule, and NASA would like to see several launch successes before entrusting its one-of-a-kind Europa spacecraft to a new vehicle.

Also running behind schedule is the Department of Energy's effort to develop an advanced version of the plutonium-fuelled batteries that powered the Voyager, Galileo, and Cassini spacecraft. These batteries are essential for any mission to the outer Solar System. The department recently abandoned one battery design for another, and is not expected to have a working version ready before the 2006 launch opportunity.

Even if these problems are solved, the Europa mission will still be "very, very challenging, about the hardest thing we're trying to do," says Richard Terrile, chief scientist for JPL's outer planets programme. The radiation environment at Europa will be fierce, worse than a satellite in Earth's ionosphere would endure following an all-out nuclear

war. The orbiter's productive lifetime will be limited to about a month as a result.

If the Europa launch slips, the Pluto-Kuiper Express, a fly-by of the Solar System's last unexplored planet, would presumably move ahead in the queue to launch in December 2004 as the first in NASA's new line of outer planets missions. Unlike the Europa orbiter, the Pluto mission has a time constraint. If it waits until 2006, the less favourable planetary alignment nearly doubles its eight-year flight time.

The proposed delay to the Europa mission is all the more disappointing now that scientists are gaining confidence that an ocean really does exist there. The Europa orbiter would probe the suspected ocean with an ice-penetrating radar. Where there is liquid water there may also be life, and a National Research Council panel concluded recently that Europa was as worthy a scientific target as Mars for spacecraft exploration. With the proposed slip, the orbiter data will not come back until 2009 at the earliest.

Nor is it certain that the space agency can afford to launch both the Pluto and Europa missions within two years of each other. Following the loss of two Mars spacecraft, NASA is coldly assessing the technical and budgetary risks of all its projects, and reconsidering what it can afford. "I think we're all taking a step back," says Terrile. Before the recent reassessment, the Europa mission had been "on the path to making similar types of errors" as the doomed Mars project, he says.

Ken Nealon, JPL project scientist for the planned Mars Sample Return mission, said last week that although nothing has been decided, the first launch of a sample return mission will probably not happen until late 2008 — six years later than what was planned only a few months ago.

Tony Reichhardt

♦ http://www.jpl.nasa.gov/ice_fire

Spain gives science and technology a ministry of its own

Barcelona

Spain's re-elected prime minister, José María Aznar, announced the creation of a new Ministry of Science and Technology last week. He said that this was a response to the need for research, innovation and the 'technological revolution' to be at the heart of the new government.

The ministry will be responsible for policy on basic and applied research, information technology and telecommunications. It will also gather those research activities that had been spread among other ministries, and will incorporate the country's Higher Council of Scientific Research.

César Nombela, the council's president, says: "the creation of the Ministry of Science and Technology is great news for the Spanish scientific community."

The ministry's policy on basic and applied research will follow the strategic 'action lines' set out in the National Plan on Research, Development and Technological Innovation 2000–04 (see *Nature* 402, 223–224; 1999). The government intends to use this plan to identify research areas in which Spain can be more competitive.

The new minister is Anna Birulés, an economist and, since 1997, director-general of the telecommunications giant Retevisión. She was previously vice-president in the Department of Industry of Catalonia's regional government.

The Ministry of Education and Culture, which had been responsible for science and higher education, will be led by Pilar del Castillo, president of the government's Centre for Sociological Investigation since 1996.

The government believes that Spain's social and economic situation makes this the right moment to attempt to raise support for research to the level of most developed countries. Achieving this goal will require close collaboration with the private sector: one of the science ministry's first tasks will be to create a biomedical research fund with the pharmaceutical industry.

In his investiture speech to parliament on 25 April, Aznar said: "[Spain] has to end the traditional weakness of its scientific and technological system". He added that the government wants the proportion of gross national product spent on research and development to rise from 0.9 per cent to 2 per cent by the end of the parliament.

Xavier Bosch

New panel set to beat biotech's bad image...

Munich

Public concern over the social impact of the life sciences has prompted the European Commission to set up a panel of 11 prominent biologists to advise it on the scientific aspects of social controversies about biotechnology and biotechnology.

The 'Biosciences High Level Group' (BHLG), announced last week, has been established by research commissioner Philippe Busquin to complement the commission's existing advisory group on bioethics.

The move aims to establish direct contacts between life scientists and European decision-makers on a more regular basis. The lack of active lobbying by scientists — compared with big companies and pressure groups — is seen by some as an important reason for their relatively small influence in the European Commission's policies (see *Nature* **398**, 646; 1999).

The BHLG will also campaign for more effective industrial exploitation of new biotechnological tools in areas such as safe food production, decomposition of harmful substances, and the preservation of biodiversity. At the same time, it will suggest how new public controversies around the life sciences might be avoided by improved information and education.

The panel features scientists from nine European countries, including biochemist Sir Tom Blundell, the former chief executive of the UK Biotechnology and Biological Sciences Research Council, Nobel laureate Rolf Zinkernagel, of the University of Zurich, and the eminent German geneticist Ernst-Ludwig Winnacker.

The group's members have been chosen for both their scientific standing and their ability to communicate with lay people. Winnacker, for example, has long worked for the public understanding of genetics in Germany, and Zinkernagel campaigned against proposed restrictions on the use of genetic engineering techniques in Switzerland (see *Nature* **388**, 315; 1997).

Axel Kahn, a geneticist at the Institut Cochin in Paris, will chair the panel. According to Kahn, the group's primary goal is to suggest ways of bridging the widening gap between the scientific and commercial opportunities presented by the biosciences and European citizens' concerns about issues such as genetically modified food, genetic screening, biomedicine and the patenting of biotechnological inventions.

Indeed, the preliminary results of a 'Eurobarometer' opinion poll on biotechnology, published last week, suggest continuing insecurity and ignorance about these issues (see right). The public's trust in political expertise also seems to be in decline.



Public distrust: rebuilding confidence in biotechnology needs help from scientists.

"I want to ask scientists to return to the debating table, as Europe needs to make sure that it has a sound basis for discussing these issues," Busquin said after the advisory groups inaugural meeting last week. He added that a "conscious political decision" relating to the biosciences is not possible without informed advice and public debate.

The group is not linked to the commission's Framework programmes on research, nor is it supposed to make direct funding recommendations. Rather it will act as a think-tank, developing measures for reconciling the interests of the public, industry and basic researchers in Europe. The practical details of how the group will approach

these goals are yet to be worked out.

The BHLG's first task will be to advise the commission on the topics and purposes of a summit on the biosciences, planned to take place in Brussels in November. The meeting will give the major European stakeholders in biotechnology, including the public, a platform for their views.

"Many Europeans believe that scientists are trying to force things on society which people neither want nor need," says BHLG member Marc Van Montagu, a plant geneticist at the University of Ghent and founder of the Ghent-based company Plant Genetic Systems.

"If the next generation of Europeans is to be more confident in biotechnology, we need to show that research in the life sciences is not being carried out just for the benefit of a handful of companies, scientists and brokers, but that it serves the whole community."

In contrast to other European research advisory bodies, such as the former European Science and Technology Assembly, which tend to include both academic and industrial researchers, the new group will take a purely science-based perspective.

According to Van Montagu, this will help it meet its goals. "The European Commission is often perceived by the public as being industry-biased," he says. "This has somewhat tarnished the image of commission-funded research, particularly in the life sciences".

Quirin Schiermeier

... as European public remains sceptical

Paris

Public scepticism in Europe towards biotechnology has increased in Europe over the past three years. But at the same time, public confidence in all sources of information about biotechnology — including both environmentalist and academic groups — has fallen, according to an opinion survey published last week.

The poll, taken last year as part of the regular Eurobarometer exercise funded by the European Commission, surveyed 16,000 people in the 15 member states of the European Union. It assessed their attitudes towards new developments such as information technology, telecommunications, space exploration and new materials.

While the approval levels of these categories have stayed

stable since the previous poll three years ago, biotechnology suffered a significant drop: 41 per cent of those polled said that biotechnology would improve the quality of life in the next 20 years, compared with 47 per cent in 1997. Only nuclear technology ranked lower on the confidence scale, with 26 per cent support.

Some researchers, however, were surprised that the drop in confidence was not greater. "Considering the scale of the debate on GMOs, we were expecting Europeans to show even more distrust now than three years ago," says Daniel Boy, director of research at the Centre d'étude de la vie politique française, one of the groups in the study.

Opinions vary widely on the applications of biotechnology. Most of those polled favoured

genetic screening for inherited diseases, the use of GM organisms to clean up pollution, and of genetic engineering to develop medicines and vaccines. But the production of GM foods, cloning human cells or tissue to treat patients, and cloning animals for medical applications received much less support.

One of the most marked changes since 1997, according to the poll, is the drop in public trust in professional organizations. Only 17 per cent see international institutions as reliable sources of information, and just 15 per cent view national public authorities as dependable.

Eurobarometer also found that respondents feel inadequately informed about biotechnology — 81 per cent feel uninformed on these issues.

Heather McCabe

Battle lines shift in stem-cell funding fight...

Washington

A new strategy was introduced last week into the stormy debate on whether the US government should support research on human stem cells. A move was made to draw a sharp line between research on the embryonic and adult versions of the pluripotent cells.

There is broad agreement that stem cells could provide a range of valuable replacement parts, from insulin-producing islet cells to treat juvenile diabetes, to neural cells for treating neurodegenerative disorders such as amyotrophic lateral sclerosis (ALS), otherwise known as Lou Gehrig's disease.

But there is disagreement over the funding of stem-cell research. Some scientists and activists say that ethical quandary could be avoided by funding research using adult stem cells, but not embryonic stem cells. But others argue that limiting research to adult cells could prevent the field's therapeutic promise from being fully realized.

The ethical contention exists because deriving stem cells from embryos destroys them. The scientific conflict arises because many researchers think embryonic stem cells could be more useful than their adult counterparts. Embryonic stem cells can potentially



Reeve: told the committee that stem cells from discarded embryos should be used for research.

become many more types of cells, and may be able to divide for longer than adult stem cells.

Both ethical and scientific reasons for and against federal funding for embryonic stem-cell research were aired at a Senate committee hearing last week that may have been a dress rehearsal for what Arlen Specter (Republican, Pennsylvania) promises to be "a knock-down, drag-out" fight in the Senate.

Specter and Tom Harkin (Democrat, Iowa) are backing a bill that would allow federal funding for both the derivation and use of embryonic stem cells. The bill, which Specter says will reach the Senate within the next month, represents just one battle in a war over stem-cell research. The US National Institutes of Health will soon issue revised guidelines that allow federal funding for the use, but not the derivation, of embryonic stem cells.

Meanwhile, private companies continue to conduct embryonic stem-cell research, unregulated by the government. And non-profit organizations may soon increase their role in both the funding and distribution of stem cells (see below).

At last week's hearing, individuals who could benefit directly from embryonic stem-cell research testified both for and against it. Mary Jane Owen, the executive director of the National Catholic Office for Persons with Disabilities, said she would rather not destroy an embryo to fix her damaged spinal cord. "We have alternatives," she said.

But the US actor Christopher Reeve, who was paralysed in an equestrian accident, said scientists should keep all avenues open. Reeve emphasized that excess embryos have long been discarded after *in vitro* fertilization attempts have ended. "Why has the use of discarded embryos for research suddenly become an issue?" Reeve asked.

Lawrence Goldstein, professor of cellular and molecular medicine at the University of California, emphasized the need to study many kinds of stem cells. Embryonic stem cells cultured by James Thomson at the University of Wisconsin, Madison, differ substantially from embryonic germ cells grown by John Gearhart at Johns Hopkins University, even though Gearhart's cells are only slightly more developed than Thomson's.

Even two different embryonic stem cell lines may behave differently, Goldstein said after the hearing. He added that allowing funding for deriving the cells is important because different extraction and culturing techniques could result in similar-seeming cells acting differently. "We probably need a bunch of different cell lines derived in a bunch of different ways," he said.

Having a wider variety of cell lines promotes better clinical science, says Jeff Rothstein, a neurology researcher at Johns Hopkins. But he adds that it will be important to fully characterize each cell line.

Rothstein is using stem cells from Gearhart, as well as neural stem cells derived from a fetus, to study cell therapy in ALS mouse models. He will soon examine adult progenitor cells in the same model. "We believe that only a comprehensive approach will be the best initial approach for a disease like ALS," he says.

Paul Smaglik

... as non-profit companies enter the fray

Washington

Industry and private foundations are taking the lead in stem-cell research, while Congress wrestles with legislation and researchers wait for revised guidelines from the US National Institutes of Health. But even though they don't face the tighter regulation and heightened scrutiny of federally funded research, private stem-cell backers are still finding that their efforts attract controversy.

The University of Wisconsin's technology transfer office, the Wisconsin Alumni Research Foundation (WARF), has set up a non-profit company to distribute embryonic stem cells isolated and grown by one of the university's researchers, James Thomson. His work was funded by a biotech company called Geron, based in Menlo Park, California.

According to WARF's managing director, Carl Gulbrandsen, the non-profit company, WiCell Research Institution, will begin distributing the cells in June,

whether or not the NIH guidelines have been completed. If they have not been completed, he says, privately funded investigators will still have access to Thomson's embryonic stem cells, whereas publicly funded scientists will have to wait.

But the terms under which WiCell is making the cells available are likely to stir controversy — the company's material transfer agreement gives it the right to require that any cells not used for the purposes expressed in researchers' applications are destroyed.

The Wisconsin state legislature may also have an impact on the company. A bill to ban the sale or transfer of human tissue narrowly missed being put up for a vote during the most recent legislative term. The bill may re-emerge in the autumn.

Meanwhile, Project ALS, a New York-based non-profit institution, is also funding stem-cell research. So far, the research it pays for has

involved only cells derived from aborted fetuses rather than from discarded embryos. Federal sources can now fund fetal tissue research, but not embryonic cell research.

The cells distributed by Project ALS have until now come from Layton BioScience of Sunnyvale, California. The company has exclusive licence on any commercial application that may result from research using its cell lines.

The Howard Hughes Medical Institute (HHMI) has formally begun considering funding stem-cell research. In a meeting last month, the institute's leadership began to weigh the pros and cons of moving ahead, but has not yet come to a decision. "HHMI will consider in a deliberate manner how it ought to proceed in this potentially vital scientific research area, recognizing both the promise of the research and the ethical challenges it presents," says spokesman Robert Potter.

P. S.

CJD survey offers Britain a glimmer of hope

London

British health officials are breathing a sigh of relief — for now. Initial results from a survey to assess the extent of infection with the human form of bovine spongiform encephalopathy (BSE) have drawn a blank.

Analyses of more than 3,000 clinical samples, mostly of appendixes removed from patients in the late 1990s, have failed to reveal any signs of the prion protein responsible for new variant Creutzfeldt–Jakob disease (vCJD). Just a handful of positive results would have suggested an epidemic of tens, or even hundreds, of thousands of cases.

But it's a case of no news rather than good news. A sizeable epidemic cannot be ruled out on the basis of these results alone. Technical limitations of the survey, plus uncertainty about the progression of vCJD in infected people, make it difficult to make predictions.

Researchers at the National CJD Surveillance Unit in Edinburgh and Derriford Hospital in Plymouth are analysing some 18,000 appendix and tonsil samples, using immunohistochemical methods to test for prions.

They hope to gauge the extent of vCJD infection in the British population. But while previous research on tissues from vCJD victims has shown that the prion accumulates in appendixes and tonsils, no one knows how early in the course of infection it appears.

Leszek Borysiewicz of the University of Wales College of Medicine in Cardiff, who chairs the scientific committee overseeing the studies, stresses the limitations of the study. "This sample size is small," he says. The tissue was also preserved with fixative chemicals that could make traces of prion hard to spot.

John Collinge of Imperial College, London, is now conducting a study of some 2,000 freshly collected tonsil samples, and will use a western blot test to detect the vCJD prion. This may yield better information. But by the time these results are available, researchers led by Roy Anderson at the Wellcome Trust Centre for the Epidemiology of Infectious Disease in Oxford claim they may already have defined an upper limit for the eventual size of the vCJD epidemic.

In January, the team published a paper in



Organ survey: the appendix, at the bottom of this false-colour X-ray, provides a test for vCJD.

Proceedings of the Royal Society B (267, 23; 2000) predicting that, if the rate at which new cases of vCJD appear remains at its current level until the end of this year, then the upper limit for the epidemic would be 14,000 cases. So far, at least, there are no signs of an increase in the rate of new cases.

Peter Aldhous

Space-station airlock to serve as temporary lab

Munich

Forget purpose-built laboratories, all you need to do experiments in space is an airlock. At least, that is where the first experiments on board the International Space Station (ISS) will take place.

Thanks to a deal with the Russian Space Agency, German and Russian physicists will make an early start to research on the space station, long before its main laboratories reach orbit. Using a makeshift lab, the researchers will begin a multimillion dollar investigation into the properties of a complex plasma later this year.

The location will be the bubble-shaped connection chamber between the Russian service module, due to be launched in June, and the Zarya control module, already in orbit. Once the ISS is fully operational, the chamber will be a busy thoroughfare, and will have to be kept clear. But until then, the

Russian Space Agency is happy to turn the airlock over to the plasma physicists, led by Greg Morfill of the Max Planck Institute for Extraterrestrial Physics in Garching.

Accustomed to a make-do-and-mend approach from their experiences with the ageing Mir station, Russian space officials have even allowed the scientists to drill a hole through a door hatch in order to thread electricity cables into the airlock. "The agency was very fast and unbureaucratic," says Morfill, who acknowledges that its western counterparts would not have been quite so accommodating. "The same sort of agreement with NASA would have required a square metre of paperwork; with

hole through a door hatch in order to thread electricity cables into the airlock. "The agency was very fast and unbureaucratic," says Morfill, who acknowledges that its western counterparts would not have been quite so accommodating. "The same sort of agreement with NASA would have required a square metre of paperwork; with

space agencies, perhaps a square kilometre," he says. The experiments will examine the unusual plasma states that can be created by introducing microspheres of melanine formaldehyde into normal plasmas. The microspheres, which are around one hundred billion times more massive than the electrons and ions that constitute a plasma, introduce order into what is normally the most disordered form of

matter. The plasma liquefies, and then crystallizes. However, gravity disturbs the spheres making it difficult to interpret data from studies of the plasma on Earth.

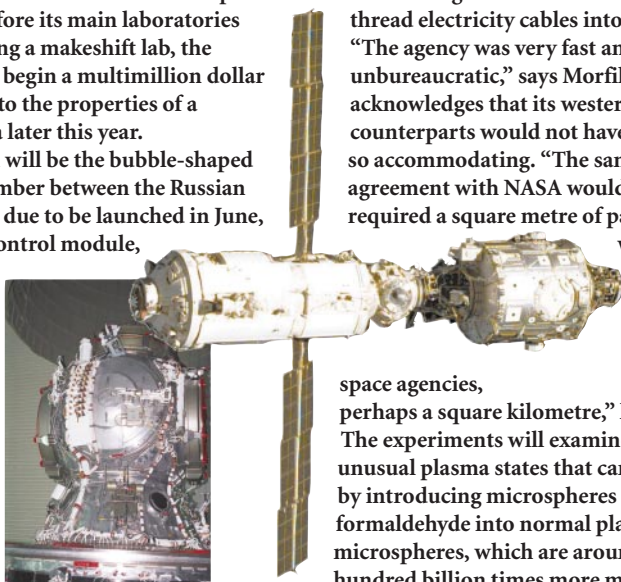
Morfill and his colleagues had originally hoped to run the experiments on Mir, but when its future started to look shaky, they shifted their allegiance to the ISS. The team will attempt to observe the propagation of waves through the crystals, and will also watch the behaviour of individual microspheres as the crystals melt — analogous to studying individual molecules in a conventional crystal. "Because the microspheres are so relatively massive," says Morfill, "we will be able to record their individual movements with [charge-coupled device] cameras."

Russia is paying for the launch costs, the crew and their training, and other logistics, to a value estimated at around DM20 million (US\$9.45 million). Germany is contributing DM8 million.

The first crew to visit the space station in October will tend the apparatus for a week, and bring back 40 hours of videos for Morfill's team to analyse. Based on these results, the experiments will be redesigned and repeated by the third ISS crew a couple of months later. The fifth crew may also continue the experiments.

Alison Abbott

♦ http://www.mpe.mpg.de/www_th/plasma-crystal/index.html



Lab space: the airlock (above) will connect the next module to the space station (right).

NSF celebrates half a century with big plans

Washington

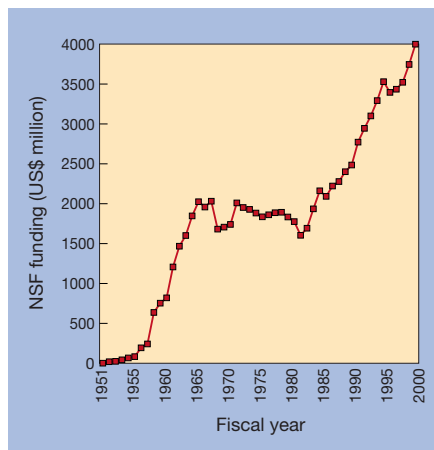
As the National Science Foundation (NSF) approaches its fiftieth birthday next week, its prospects have seldom looked brighter.

The Clinton administration has proposed a \$675 million budget increase next year for the agency, which funds most non-biomedical research at US universities. This is by far the largest increase in its history, and there are frequent statements of support for it in the US Congress.

The NSF has been asked to lead most of the government's recent science initiatives, such as those in computing and in nanotechnology. Rita Colwell, the marine biologist who became NSF director in 1998, talks publicly of raising its existing budget of \$4 billion to \$8 billion, or more. These are large numbers for an agency which, although highly respected around the world for the quality of its work, has served for most of its history as a relatively small component of the US government's \$80 billion annual research and development effort.

The National Science Foundation Act was passed in 1950, after five years of wrangling about how the US government should continue the large-scale support for research begun during the Second World War. The delay meant that the NSF started small, supporting academic research in those disciplines that other areas of government — such as the National Institutes of Health, the Naval Research Laboratory and the Atomic Energy Commission — had not grabbed for themselves in the immediate post-war period.

The Soviet Union's launch of Sputnik in 1957 inspired an upsurge in the agency's fortunes (see graph). But its budget then remained essentially flat until the early 1980s, when economists and politicians began to associate basic research with economic growth. This association has gained acceptance during the recent economic boom.



On the up: recent political enthusiasm for basic research has been good for the NSF's budget.



Colwell: NSF director is "guardedly optimistic" that Congress will grant a record budget request.

"For the first time, there's a growing recognition of the impact of basic research on the quality of life, and the impact of the NSF on basic research," says Eamon Kelly, an economist at Tulane University in Louisiana and chairman of the National Science Board, the presidentially appointed committee which supervises the NSF.

"Basic research wasn't so valued during the Cold War years," agrees Howard Silver, chairman of the Coalition for National Science Funding, which lobbies for the NSF in Washington. "People are now making connections between basic research and prosperity, better medicine and good times."

The NSF is seen by the administration, and by members of Congress who are interested in science and technology, as an efficient funding source of such basic research. It has no laboratories of its own, but, despite the bureaucracy of its headquarters in an office building in northern Virginia, it maintains close ties with the academic community.

About one third of its programme managers, who vet grant applications and arrange for their review, are academics temporarily seconded to the agency. "It's the closest thing you get to a university environment inside the government," says George Mazuzan, who recently retired as the NSF's historian. "There's a profusion of outside influence that other agency's don't get."

The agency has always been characterized by a competent, low-key management style. "By and large it has had good managers — there have been no real losers," says Frank Press, former president of the US National Academy of Sciences and science adviser to President Jimmy Carter. "And the politicians have more or less stayed away from it."

But this mild-mannered approach has its limitations, observers point out. The academics who have led the agency "are not very politically acute people," says Dan Greenberg, a columnist for *The Washington Post*

and visiting scholar at Johns Hopkins University in Baltimore.

Since her appointment, Colwell has been trying to change that. From the outset, she told sometimes disbelieving audiences that the National Science Foundation could become an \$8 billion agency. Her tone contrasted sharply with the always-cautious public statements of her predecessor, Neal Lane (who left to become Clinton's science adviser), and surprised some long-time observers. Now, however, "the community has grown to see that her style might have some benefits," according to Sam Rankin, associate executive director of the American Mathematical Society.

The main such benefit has been the administration's unprecedented request for a \$675 million budget hike, which, as Colwell pointed out at an NSF birthday event last week, is twice as large in dollar terms as any previous budget request in the agency's history. Congress appropriations committees will now decide how much of this should be granted.

"Very strong support has been expressed for us because of the work we do for the country," says Colwell. "We are guardedly optimistic."

But as Colwell works the marble halls of Capitol Hill this spring to secure backing for the increase, she faces the same problem as her predecessors. The agency is widely liked, but it lacks any individual champion who will defend it to the death. Senator Barbara Mikulski (Democrat, Maryland), for example, will protect the NASA budget, which secures thousands of jobs at the Goddard Space Center, to her last breath — and most of the programmes that compete with NSF for funds have similar protection. As Rankin puts it: "No-one will go to the wall for NSF."

The NSF is working to broaden its appeal. Its anniversary celebrations, starting with last week's event at which Colwell and Nobel laureate Leon Lederman taught Washington schoolchildren how to do science, will emphasize the foundation's school education programmes. "Science education is a serious business, but it's also great fun," Lederman told the children.

Colwell invited accomplished scientists to join the 240 already taking part in an anniversary programme introducing children aged between 12 and 14 to science. In addition, a competition held in the Sunday newspaper magazine *Parade* asks children to submit their ideas for applying science to "national and global problems".

Kelly says that the NSF made a "conscious decision" to focus the anniversary celebrations on the public. "We've not done a good job of communicating what the agency does to the general public," he says. **Colin Macilwain**

French success with immune disorder boosts gene therapy

Paris Two infants with inherited severe combined immunodeficiency (SCID) have been treated with gene therapy. The positive findings will boost a field still shrouded in gloom after the death of Jesse Gelsinger, who was enrolled on a gene therapy trial at the University of Pennsylvania, last year.

The form of SCID in question is caused by a mutation in a gene on the X chromosome that codes for a subunit of certain receptors for molecules known as cytokines. Patients suffer from defects in two classes of immune cells: T cells and natural killer cells.

A team in Paris, led by researchers at the French medical research agency INSERM, administered the sequence for the receptor subunit using a vector derived from a retrovirus. In *Science* (288, 669; 2000), the researchers report "full correction of disease phenotype" ten months after the treatment.

Public invited to suggest where to point Hubble

Washington In celebration of the Hubble Space Telescope's tenth anniversary, the US space agency NASA is inviting the public to pick targets for the telescope. Hubble's operators will take suggestions up to 6 June on where the telescope should be pointed.

The selection criteria include feasibility, scientific interest and the potential to take a compelling image. The target must not be something that Hubble has observed before.

This is the third such opportunity offered to the public. The first two winners voted for the Polar Ring Galaxy NGC 4650A and a group of galaxies called Hickson Compact Group 87. Guidelines to help in the selection can be found at: <http://heritage.stsci.edu>.

US treats AIDS as a threat to security

Washington The Clinton administration has declared AIDS a potential threat to US national security. As a result, the administration will seek to double funds — to \$254 million — in the fiscal year 2001 for combating the disease in other countries. The National Security Council, which normally deals with threats of war and terrorism, will be given responsibility for drawing up an international AIDS strategy.

The new emphasis on AIDS outside the United States was triggered in part by an intelligence report last year, which concluded that the disease could cause political instability in Africa, Asia and the countries of the former Soviet Union. "Hopefully this move will underscore the need to view AIDS more as an international disease," says

Anthony Fauci, director of the National Institute of Allergy and Infectious Diseases at the National Institutes of Health.

Russian Supreme Court clears navy officer of treason

Moscow Russia's Supreme Court has upheld the acquittal by a St Petersburg court of Alexander Nikitin, a retired Russian navy captain, on espionage and treason charges. Nikitin, now an environmental campaigner, co-authored a report critical of the Russian navy's handling of nuclear wastes.

The report led the Federal Security Services to pursue espionage charges over a five-year period. Amnesty International welcomed the decision, although it says that the ruling could be appealed again. "There never was any doubt that Alexander Nikitin was innocent," says a spokesperson for the organization.

Life isn't sweet for young UK academics

London Young academics in England feel underpaid, under pressure and undervalued, according to a survey for the Higher Education Funding Council for England, which funds teaching and research in English universities. Most of the 370 interviewees were dissatisfied with the succession of short-term contracts that currently comprises the career structure for young academics, and felt that attempts to improve the situation had failed.

The survey's finding on pressure in the workplace matches recent research presented to the British Psychological Society, which found that more than a quarter of academic staff surveyed had suffered from stress-related illnesses in the previous year.

Claims on matter of gravity receive cautious welcome

London Physicists from the University of Washington in Seattle claim to have narrowed the uncertainty surrounding the value of G , the gravitational constant — a value that has fluctuated disconcertingly over the past few years. Because of gravity's relative weakness, G is one of the hardest fundamental constants to measure.

The Washington team, led by Jens Gundlach and Stephen Merkowitz, claim to have established the value to within 0.0015 per cent, using a modified version of the 'torsion balance technique' first used to make the measurement in 1798.

But Clive Speake at the University of Birmingham, UK, who will soon announce further measurements, says that the new figure needs confirmation using different techniques. Speake's team is holding back its results until it has two independent measurements that tally.



Hand transplant in danger of rejection

Paris The world's first hand transplant patient is being treated for rejection symptoms, the Sir Charles Gairdner Hospital in Perth, Australia, confirmed last weekend. Clint Hallam (above) received the transplant in September 1998 at the Hôpital Edouard-Herriot in Lyon, France. One of the questions has been whether the novel cocktail of antirejection drugs that made the operation possible would be effective in the long term.

Jean-Michel Dubernard, who carried out the operation, was unavailable for comment as *Nature* went to press. But he is reported to be confident that Hallam's hand can be saved. An earlier bout of rejection was controlled, and tests on Hallam carried out last month had shown low levels of the immunosuppressive drugs, suggesting that the treatment regime needed to be adapted.

US agricultural research 'at risk', says report

Washington The US Department of Agriculture's main basic research programme needs reorganization and a fourfold increase in funding — to \$500 million a year — according to a study from the National Research Council, the research arm of the National Academy of Sciences.

The study says that inadequate funding has placed the programme, known as the National Research Initiative, "at risk". It calls for larger grants, a new external advisory panel, and an end to a ceiling on overhead charges that deters some universities from applying to the programme.

The National Research Initiative was set up in 1991 to support more competitive grants in basic agricultural research. But its budget has never come close to the \$500 million originally envisaged. Most of the Department of Agriculture's \$1.6 billion research and development programme supports non-competitive grants and contracts.

The missionary from Munich

Germany's network of national research centres is trying to reinvent itself. Alison Abbott talks to Rudi Balling, the dynamic biologist charged with revitalizing one of the country's scientific underachievers.

Any research institute with the word 'national' in its title should be something special. Especially if it is in Germany, a country with a reputation for scientific excellence. But curiously many of Germany's 16 national research centres are failing to live up to this billing.

Ask researchers from outside the country to identify a leading German research centre and the chances are they will name one of the prestigious institutes run by the Max Planck Society. It is unlikely, however, that anyone will mention the National Centre for Biotechnological Research (GBF) in Braunschweig.

But biologist Rudi Balling has plans to change all that when he takes over the reins at the GBF later this year. At 47, Balling will be the youngest director a German national research centre has ever had. This is no coin-

cidence — Balling's appointment is part of a broader push by all of the centres to rejuvenate and modernize themselves. "We very much need young directors," says Detlev Ganten, director of the Max Delbrück Centre for Molecular Medicine in Berlin, and chairman of the Helmholtz Association, the umbrella organization that oversees the national research centres.

Reversing the decline

Germany's first national research centres were established in the mid-1950s to support nuclear research (see table). Later additions focused on other specific missions, mostly in applied science. But as political opinion swung against all things nuclear in the 1980s, the nuclear centres shifted their efforts towards environmental research. In doing so, they lost focus. Some other cen-

More than ever we have to justify our existence, and you need a strong mission to do this, which also has to be unique.

tres, including the GBF, failed to keep pace with developments in their fields.

German unification brought a flurry of activity, as three new centres were created in the former East Germany — including the Max Delbrück Centre. But expansion eastwards did not solve the fundamental problems. Although there are islands of excellence — the German Cancer Research Centre (DKFZ) in Heidelberg, for instance, boasts leading groups in molecular biology, and the DESY in Hamburg remains one of the world's premier centres for high-energy physics — the general story has been one of slow decline.

Over the past decade, the staffing budgets for all the national research centres have been cut by one-and-a-half per cent each year. Politicians have accused many of the centres, with their ageing populations of tenured research staff, of being out of tune with the needs of modern German society. In this context, Balling's appointment is significant, because he has already initiated a turnaround at the National Research Centre for Environment and Health (GSF) in Munich, in his current role as director of its Institute of Mammalian Genetics.

Spend a few hours in Balling's company, and the impression that comes across is of a scientific version of the film star Robin Williams. He is quick-talking, humorous and energetic. Balling also has the knack of being in the right place at the right time. After completing a PhD in nutrition at the University of Bonn in 1985, he moved into developmental biology. In 1987, he joined Peter Gruss, director of the Max Planck Institute for Biophysical Chemistry in Göttingen, who was the first to clone members of two key families of development genes, *Hox* and *Pax*, in the mouse.

Balling's job was to find out what the genes controlled, and he became the first to assign a definite function to one of these genes in a vertebrate. He showed that

Germany's national research centres					
Name	Abbreviation	Founded	No. of staff	Location	Research areas
Research Centre Jülich	FZJ	1956	4,200	Jülich	Materials, energy, informatics, biotechnology, environment
Research Centre Karlsruhe	FZK	1956	3,508	Karlsruhe	Physics, energy, biotechnology, environment, microsystems engineering
GKSS Research Centre Geesthacht	GKSS	1956	750	Geesthacht	Environment, materials
Deutsches Elektronen-Synchrotron	DESY	1959	1,550	Hamburg	High-energy physics
Hahn Meitner Institute	HMI	1959	678	Berlin	Materials, solar energy
GSF — National Research Centre for Environment and Health	GSF	1960	1,500	Munich	Health, environment
Institute of Plasma Physics	IPP	1960	980	Garching	Plasma physics
German Cancer Research Centre	DKFZ	1964	1,594	Heidelberg	Cancer research
National Centre for Biotechnological Research	GBF	1968	575	Braunschweig	Biotechnology, health, environment
German National Research Centre for Information Technology	GMD	1968	1,200	Bonn	Informatics
German Aerospace Centre	DLR	1969	4,500	Cologne	Space, aeronautics
National Research Centre for Heavy-Ion Research	GSI	1969	700	Darmstadt	Heavy-ion research
Alfred Wegener Institute for Polar and Ocean Research	AWI	1980	620	Bremerhaven	Polar and oceanic research
UFZ Centre for Environmental Research Leipzig-Halle	UFZ	1991	620	Leipzig	Environment
Max Delbrück Centre for Molecular Medicine	MDC	1992	700	Berlin	Medical research
GeoForschungsZentrum Potsdam	GFZ	1992	585	Potsdam	Earth sciences



ANTJE LANGER

Balling: sitting on a scientific "gold mine".

overexpression of the *Hoxa-7* gene leads to a change in the characteristics of vertebrae in mice. "It made people aware of the power of mouse genetics," he says. Later, Balling's work on a mouse *Pax* gene led to the first association of the gene family with a human inherited disorder: Waardenburg syndrome, which is characterized by skull abnormalities, often accompanied by deafness.

His big break came in 1993, with the post at the GSF. At that time, the pressure on the centre to justify its existence with a credible mission was at its height. But Balling arrived with a clear concept of how to modernize his department.

"I found strange radiation labs, with a bit of genetics, and big old animal houses with well-trained and motivated staff," Balling recalls. This set-up was designed to serve the GSF's original mission of investigating the biological effects of radiation — "old-fashioned risk assessment", Balling calls it. But he realized it could also serve "modern-style risk assessment" — the study of genetic disposition. He encouraged the GSF to refocus its efforts onto the genetic influences underlying the varied ways individuals respond to drugs and toxic chemicals — the emerging fields of pharmacogenomics and toxicogenomics. He added programmes in developmental genetics and human genetics, and launched a genome analysis centre.

One of Balling's most significant achievements is his conversion of part of the GSF's animal facilities into a large-scale mutagenesis screen to support research into gene function. This screen — one of only two in the world on such a scale — uses chemical mutagens randomly to generate mutant mice. Tens of thousands of mice are screened every year. By correlating their particular abnormalities with the genes that are mutated, scientists can investigate what these genes normally do. The screen is supported by the German Human Genome Project, which was launched in 1995, just at the right time for Balling to begin his project on the scale he felt was needed.

Not surprisingly, given the proliferation of new projects, Balling's department has increased from 30 scientists to more than a hundred. "I wanted to expand as fast as possible, because to my mind critical mass is everything," he says.

Focusing on the future

Balling now wants to bring a similar approach to the GBF. "Things have to change," he explains. "More than ever we have to justify our existence, and you need a strong mission to do this, which also has to be unique."

The GBF was founded in 1968, primarily to develop biological processing technologies. But as the cutting edge of biotech research has moved away from the big fermenters that were the GBF's forte, the centre has found itself in need of a fresh focus. Balling believes he can provide one.

While many of the world's biologists prepare for the post-genomic era, in which the main task will be unravelling the functions of newly sequenced genes, Balling's vision for the GBF is to surf the next wave — genome-genome interactions. Specifically, Balling plans to develop an interdisciplinary approach to host-pathogen interactions. He wants to forge links between the GBF's currently disjointed departments, using genetics and genomics as the common ground. He aims to address questions about individual susceptibility to both disease and antibiotic resistance, and to investigate what happens when two genomes — for example, from a bacterium and a human — meet and fight it out.

The potential is enormous, Balling claims. The GBF already has many of the right tools, including sequencing and proteomics facilities, courtesy of the German Human Genome Project. A brand new mouse facility has opened recently, and the GBF shares a campus with the German Collection of Microorganisms and Cell Cultures, and one of Germany's only academic facilities certified to prepare proteins or DNA for clinical use. "It's like sitting on a gold mine," says Balling.

Critical to the success of Balling's initiative, however, is the need for more genomic

information on bacteria. Fortunately, the German Human Genome Project will next week announce a microbial genome programme. As ever, it seems, Balling finds himself in the right place at the right time.

But will the GBF's gain be the GSF's loss? Ganten thinks not, because Balling has taken care to develop the careers of the scientists he is leaving behind. "Rudi is a very generous person and is leaving many of the projects he established at the GSF in the hands of even younger colleagues," says Ganten. "The GSF will survive very well."

Rewarding excellence

The Helmholtz Association is also planning changes which it believes will strengthen the entire research centre network. It wants to concentrate resources in the hands of the most dynamic researchers across all the national centres. At present, each centre receives 90 per cent of its core research funding from the federal government, and distributes this budget among its scientific staff. Instead, the association wants to pool most of the federal money into six programmes covering areas considered to be national priorities: life sciences, energy, basic physics, transport, environment and space. Researchers from across the entire network of national research centres would then compete for these funds.

By placing the future of Germany's national research centres in the hands of scientists such as Balling, and moving funding into competitive research grants, Ganten believes the Helmholtz Association is well on the way to enacting the reforms that have been demanded. "In response to political pressure, we have changed the orientation of many of our research centres," he says.

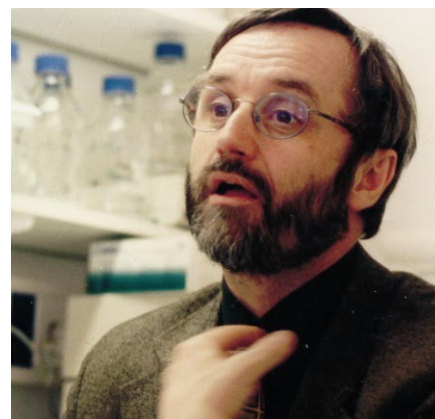
Ganten is now looking to the politicians to reward the centres for these changes. "We have been suffering too many cuts in budget for too many years," he claims. "Now we are more agile, but we really do need more money to ensure that our efforts are productive." ■

Alison Abbott is Nature's Senior European Correspondent.

► <http://www.helmholtz.de>

► <http://www.gbf.de/index-uk.html>

► <http://www.gsf.de/gsf/englisch/index.html>



ANTJE LANGER

Sagan breached security by revealing US work on a lunar bomb project

Sir — In his review of two biographies of Carl Sagan, by William Poundstone and by Keay Davidson (*Nature* **401**, 857; 1999), Christopher Chyba deals extensively with “Davidson’s accusation that the young Sagan wilfully and illegally revealed classified information...” and further states: “This is a serious and specific legal allegation which Davidson does not substantiate”.

The classified project involved, Project A119, entitled A Study of Lunar Research Flights (SECRET), was conducted at Armour Research Foundation (ARF) while I was manager of physics research. I was also leader of the project, so Sagan reported directly to me. I, therefore, feel obliged to extend the historical record beyond the Davidson biography by offering some additional, first-hand comments.

A119 was one of a series of projects conducted at ARF under my direction from 1949 to 1962, all concerned with the global environmental effects of nuclear explosions and related phenomena. Some time before May 1958, we were asked by the US Air Force to add a small, fast-track project to investigate the visibility and effects of a hypothetical nuclear explosion on the Moon. I was told the Air Force was very interested in the possibility of a surprise demonstration explosion, with all its obvious implications for public relations and the Cold War.

Whether the project was motivated by a desire for the United States to impress the world (and the Soviet Union in particular), or by fear that the Soviet Union itself might try the stunt, I cannot say. It was emphasized, however, that the most sensitive aspect of the project was, as with many other Department of Defense projects, its very existence. Hence it was given a separate name and work on it was classified as secret. Our task was to assess what gross visible and other phenomena might be generated by the explosive release on the lunar surface of a given number of kilotonnes of energy.

We were also asked to try to determine what legitimate scientific data might be gathered from such an event: for example, about lunar chemistry. The cost to science of destroying the pristine lunar environment did not seem of concern to our sponsors — but it certainly was to us, as I made clear at the time.

In staffing A119, I realised that we needed expertise in planetary physics and asked Gerard P. Kuiper to act as our consultant. Kuiper agreed, and in time suggested that I hire a graduate student from Yerkes/

University of Chicago called Carl Sagan, who needed a job. I gave Sagan the assignment of mathematically modelling the expansion of an exploding gas/dust cloud rarifying into the space around the Moon. This was preliminary to attempts to calculate the visibility of such a cloud from Earth. Sagan had difficulty with the problem and consulted Kuiper several times before I brought in additional help. Sagan soon suggested that he should try to see how a nuclear explosion might be used to detect organic molecules on the Moon. I agreed to a brief effort in that direction.

Nine monthly progress reports, all classified as secret and including all of Sagan’s work, were issued by ARF to the Air Force Special Weapons Center under Project A119 from May 1958 to January 1959. According to Armour (now the Illinois Institute of Technology Research Institute) archives, they were all destroyed after October 1987.

I did not know until the biographies were published that Sagan had sent an “unclassified” (by whom?) manuscript about his work on A119 to any unauthorized person — let alone to five people, as Chyba remarks. Both ARF’s and Sagan’s obligations under the A119 contract, as Sagan’s signed security agreement would have clearly informed him, required Air Force clearance of any such revelations. As his boss at the time, I would have had to take forward any such request, and Air Force permission would have been extremely unlikely in those long-ago, very tense times.

In his review, Chyba argues that classified documents can contain unclassified titles or subsections. Perhaps so, but that misses the point. Fortunately for the future of lunar science, a one- or two-horse race to detonate a nuclear explosion never occurred. But in my opinion Sagan breached security in March 1959 when he revealed the ARF’s classified projects on “possible lunar nuclear detonations” in his application for a Miller fellowship.

Leonard Reiffel

Exelar Corporation, 602 Deming Place, Chicago, Illinois 60614, USA

Austria takes lab animal welfare seriously

Sir — In response to the News report “Austria taken to court for inadequate laws on animal welfare” (*Nature* **403**, 582; 2000), I would like to explain the legal situation. Animal experiments in Austria are regulated on the federal level by the Animal Experiments Act. Passed in 1989, well before Austria’s accession to the European Union in 1995, this law was drawn up to conform with the Animal Experiment Directive (86/609/EEC), even

though at that time Austria was not obliged to do so. The act was amended last year. Animal welfare is regulated at provincial level, and all nine provinces have animal welfare legislation.

The Animal Experiments Act empowers the competent Federal Ministers to regulate technical matters by ministerial ordinance. Ordinances laying down specific rules (for example, covering research centres that breed animals on their own premises) are currently under consultation. As soon as they come into force, the last minor deficiencies in transposing the Directive will have been eliminated.

Wolf Frühauf

Federal Ministry of Education, Science and Culture, Minoritenplatz 5, A-1014 Vienna, Austria

How metrics could have saved UK car industry

Sir — You recently featured a report, in the “100 years ago” column, (*Nature* **404**, 27; 2000) on German iron manufacturing, written by the British consul in Amsterdam in 1900. He mentioned a uniform system of dimensions for articles of universal consumption, and a system of metric screw threads being fixed by a committee of engineers’ associations. Unless Britain joined the metric standard, the consul warned, “Germany and the continent generally will have a constantly increasing advantage over British manufacturers”. His words were prophetic — 150 of the world’s 190 export markets are now metric.

The engineers’ associations agreed on the *Système International* (SI) of metric threads at a special meeting, the Congress of Zurich, in October 1898. This was to end the confusion caused by different European countries using different systems on their railways, though they didn’t enter British consciousness until the Second World War, as BSI War Emergency Standard BS 1095:1943. The 1898 meeting also saw the birth of ISO metric standards for engineering, which entered the British standards environment in 1966 through a book published by the then Ministry of Technology.

Meanwhile, metric screw threads, standard diameters, shaft/hole fits and machining tolerances for the engineering industry had been internationalized during the 1920s and 1930s by a body called ISA, which was absorbed into ISO in 1946. The old SI metric thread with its undesirably sharp fillet radius was modernized by application of the unified profile by ISO in 1958 and standardized by all industrialized nations soon afterwards. That is why all mechanical engineering assemblies made in mainland European factories since 1959

use the same threads up to 80 mm diameter.

In order to assure the free circulation of engineering components and subassemblies in the European Common Market, the UK government — led by Harold Wilson — announced in May 1965 that British industry would abandon the inch within ten years and adopt ISO metric standards. The first attempt to produce a British engine to metric standards was already in progress at Leyland Motors with the Leyland 500 engine. However, this was marketed in continental Europe only in the Leyland National bus, whose full metrication was never completed: it incorporated four different screw-thread standards and was impossible to manage in European workshops.

British Leyland vehicles failed to attract continental dealers because the inch was already alien in mainland Europe — it had actually been outlawed in mechanical engineering in Germany during the 1930s. It remained only in water- and gas-pipes, and in certain items imported from the United States.

Inch threads had no future in Europe, but the planned conversion of British engineering factories did not occur. As a result, the UK's non-metric products were rejected by Europe as a nuisance.

It was the abandonment of Wilson's metrication programme, along with deficient marketing, that caused the stagnation of British Leyland and so many other engineering-based companies in the United Kingdom. The economic effects of Britain's failure to adopt the metric system on schedule have never been quantified. But the recent loss of the \$125 million NASA Mars Orbiter probe emphasizes the vast sums of money that can be lost through the unfamiliarity of the average UK or US technocrat with metric units.

Michael T. Knowles

11 Powlett Court, Bath BA2 6QJ, UK

Instrument's ability to do the job is NASA's priority

Sir — In your News profile article "Sky's the limit as teams bid for NASA Project" (*Nature* 403, 587; 2000), Colin Macilwain implies that NASA administrator Daniel Goldin's professed enthusiasm for particle physics detector technologies would be a factor in the selection of one of two competing instruments for the Gamma-ray Large Area Space Telescope Mission. (High-energy gamma-ray instrumentation has always drawn heavily from techniques developed for high-energy physics studies.) There is the further implication that contributions to the cost might also provide an advantage.

Although the winning team, from Stan-

ford, contains particle physicists sponsored by the Department of Energy, this had no direct bearing on the selection. It is NASA's policy to select instruments for its missions through a peer-review process that evaluates first and foremost the proposed science and the proposed instrument's ability to achieve that science. There are other factors, but these do not include the participation of other partners, either foreign or domestic, unless they can help the team achieve their science goals.

Incidentally, your News profile stated that the competing instrument was from NASA's Marshall Space Flight Center (MSFC). Although MSFC did participate, along with several other institutions, the principal investigator was from the University of Alabama in Huntsville, and the proposal was submitted through the university. The selected instrument also included a team from NASA's Goddard Space Flight Center, as well as several other institutions, both domestic and foreign.

Donald A. Kniffen

NASA Headquarters, Office of Space Science,
Research Program Management Division,
Washington DC 20546, USA

Putting marine mammals back in the mainstream

Sir — I read with interest Vaclav Smil's Millennium Essay¹, in which he describes the general importance to bioenergetics of Max Kleiber's studies on the scaling relationship of metabolic rate with body mass in animals. Across 18 orders of magnitude from unicellular organisms to whales, it seems metabolic rate is proportional to body mass raised to the power of 0.75; the so-called three-quarters rule, exemplified by the well-known mouse-to-elephant curve.

Smil gives the example of marine mammals as species whose metabolic rates lie outside this relationship because, apparently, seals and whales have basal metabolic rates (BMR) about twice as high as those of other animals their size, which illustrates their environmental specialization for thermoregulation in cold water. However, there is evidence that is inconsistent with this view².

The perception that pinnipeds (seals, sea lions, fur seals and walrus) and cetaceans (whales, dolphins and porpoises) have BMRs twice as high as similar sized animals is an idea that has been widely accepted for decades. It primarily arises from comparisons of marine with terrestrial mammal data.

In his studies, Kleiber was very specific about the conditions under which BMR measurements should be made, in a bid to reduce variance between comparisons of

basal rates in animals of different size and from different taxonomic groups. (Measurements should be made on mature animals in a post-absorptive state, non-reproductive, at thermoneutral ambient temperatures, and without abnormal activity².) For these reasons, Kleiber rejected the use of the two determinations for marine mammal BMR available to him when preparing his original paper.

Nevertheless, published BMR data for marine mammals often have not conformed to these criteria, but have been included in comparative analyses with data that do. This has led to the widely held view that marine mammals have higher BMRs and that they are therefore not 'normal' mammals.

In the analysis by Lavigne *et al.*², where data from studies on seals and whales were excluded when determinations did not fulfil Kleiber's criteria, it appears that metabolic rates of marine mammals were indistinguishable from those predicted for other mammals under similar conditions. In support of this conclusion, the best available data on minke whale (*Baleoptera acutorostrata*) metabolic rates, determined from field tracking studies and heat loss determinations³, give a value at zero swimming speed only 17 per cent greater than the BMR value predicted from Kleiber's general equation.

Clearly, these studies suggest that Kleiber's relationship applies just as well to marine mammals as to terrestrial species.

David W. Sims

Department of Zoology, University of Aberdeen,
Aberdeen AB24 2TZ, UK

1. *Nature* 403, 597 (2000).

2. Lavigne, D. M. *et al.* *Can. J. Zool.* 64, 279–284 (1986).

3. Blix, A. S. & Folkow, L. *Acta Physiol. Scand.* 153, 61–66 (1995).

Persian role in flowering of Islamic science

Sir — Giovanni Bignami in his Millennium Essay (*Nature* 404, 227; 2000) qualified two Persian thinkers, Avicenna and Omar Khayyām, as belonging to the Arabic world. Occidental writers frequently take "Muslim" to mean "Arab" and consider Islamic culture to be Arabic. Yet even 1,000 years ago the Islamic world was composed of people of quite different origins. Many of the thinkers who participated in the blossoming of science at that period were in fact Persians and not Arabs.

Saadi Khochbin

INSERM U309, Institut Albert Bonniot, 38706
La Tronche Cedex, France

Erratum The number of signatories to the letter "Distinguished scientists back Germany's DFG..." (*Nature* 404, 922; 2000) is 1,641 — not 1,164, as stated in the correspondence. *Nature* apologizes for this error.

Hell-fire and medication

Doctors delighted in the glow of phosphorus, but it killed more than it saved.

The Shocking History of Phosphorus: A Biography of the Devil's Element

by John Emsley

Macmillan: 2000. 326 pp. £12.99, \$24.95

Ryan J. Huxtable

In 1763 that quintessentially bad-tempered Scotsman, Tobias Smollett, consulted a famous doctor in Montpellier, France, by sending him an account of his condition in Latin. The poor doctor, clearly out of his depth in Latin, replied in French, and made so many errors that Smollett sent him another letter — with another fee — pointing out all the mistakes and confusions in his reply. Later, Smollett triumphantly reported meeting an Englishman who had received an identical letter from the physician, even though they had very different diseases.

As Smollett discovered, possession of a doctorate does not necessarily imply knowledge. John Emsley notes that the physician Hennig Brandt, who isolated phosphorus in Hamburg in 1669, was unable to translate the simple Latin name of a medication, but that is insufficient evidence to suggest, as Emsley does, that his title of doctor was unearned. Not surprisingly, the discovery of a luminescent, self-igniting substance led to glowing reports of its medicinal properties. An aphrodisiac effect was demonstrated in ducks (a suitable test species for a quack preparation). Products containing phosphorus were rapidly applied to the treatment of colic, asthma, tetanus, apoplexy, gout, depression, epilepsy and tuberculosis, in a testament to the continuing credulity of humans when it comes to miracle cures.

This book is part history, part economics and part chemistry. It presents a Manichaean dualist view, maintaining a kind of moral double-entry bookkeeping on phosphorus. In the one column we have the disease phossy jaw, bombs, and nerve gas, counterpoised in the other column by fertilizer, food preservative, fire-resistant pyjamas, cola drinks and rustless stays on corsets.

One of the practical applications of phosphorus was the development of the match industry. Matches were developed from earlier self-ignitable inventions ranging from phosphoric tapers (1781) to the superbly named Oxymuriated Ignitors of 1806 and the Promethean glass beads of 1828. These were displaced by friction lights, or lucifers, a favourite of young men who would ignite them and throw them at the feet of young ladies “to startle them”. The continued development of the match-making industry throughout the nineteenth century resulted



Alchemical reaction: Brandt discovered phosphorus by accident while trying to create gold.

in machines capable of turning out 14 million matches a day, and requiring only nine operators. The price of matches fell from 12 pennies per hundred in 1827 to 1200 per penny 60 years later. Matches are now so inexpensive that restaurants and other businesses routinely give boxes of them away.

The twentieth century saw progress in warfare with the development of phosphorus bombs and potent nerve gases such as sarin, soman, tabun and VX. In a modern version of the canary in the coal mine, soldiers are advised that if no birds are twittering on the battlefield, nerve gas may have been released and it is time to inject the antidote. Molotov cocktails consisting of bottles of elemental phosphorus in benzene were issued to the British public in the Second World War for use against invaders. (Perhaps the British had more reason than they knew to be grateful that the invasion never came.) The phosphorus story completed a circle in 1943, when Brandt's hometown of Hamburg was destroyed by firebombs in response to the bombing of British cities. Emsley, in one of his frequent dogmatic phrases, remarks that “dropping phosphorus bombs on civilians is never likely to happen again”. To

respond with a phrase from my childhood, tell that to the marines.

A Tide® in the affairs of men was reached in 1946 with Procter & Gamble's eponymous detergent, containing sodium tripolyphosphate. Yet whiter-than-white clothing came at an ecological cost when phosphate run-off was associated with algal blooms and eutrophication of lakes.

There are many linguistic and stylistic inadequacies in this book. However, the comment that spontaneous ignition of cows is “relatively nonexistent” is a poetic gem.

According to John Ruskin, the Pathetic Fallacy is the attribution of human feelings to nature. This book suffers from an overuse of the Pathetic Fallacy, from its titular claim to be a biography of phosphorus onward. John Milton consigned the fallen angels to a Hell “fed with ever-burning sulphur unconsum'd”, so there is in any case a stronger claimant than phosphorus to being the devil's element. Phosphorus was the unlucky thirteenth element to be discovered, and a major industrial accident occurred one Friday 13: “traditionally a day of ill-omen” as Emsley tells us. On the other hand, another occurred on 1 April, All Fools' Day.

Emsley tells us that phosphorus's "capacity for evil cursed all who tried to exploit it... Phosphorus was damned from the moment it was born... It promises cures but delivers curses... the ultimate evil... Phosphorus has plumbed the depths of evil", and so on. However, the evils associated with phosphorus — eutrophication, phosphy jaw, phosphorus bombs and nerve gases — are problems of humans, not nature. It is a pathetic fallacy in several senses to blame the rest of nature for our own nastiness, stupidity, narrow-mindedness, xenophobia and aggressiveness.

Probably none of the problems posed by phosphorus and its compounds are simple. As with other industries, match-making carried its occupational hazards, the worst of which was "phosphy jaw", in which the bones of the jaw rotted away. However, phosphy jaw may involve poor oral hygiene, malnutrition and substandard working conditions, as well as exposure to phosphorus. Similarly, to blame eutrophication on the use of phosphates ignores other organic and inorganic pollutants, acid rain, and the massive ecological disruptions associated with human consumption. To replace phosphate detergents with zeolites, or plastic McDonald's containers with cardboard, simply replaces one problem with another. As the laws of ecology have it, there is no such thing as a free meal.

Another of the laws states that everything has to go somewhere. Consumption creates crap. The phosphorus bombs dumped in the ocean at the end of the war wash up on beaches 50 years later. Emsley expresses that as modern-day hermeticism, by writing that the phosphate we eat returns as the detergent in our dishwasher, which returns as the phosphate in our cola drink. Or as Longfellow said: *All things must change To something new, to something strange; Nothing that is can pause or stay...*

Ryan J. Huxtable is in the Department of Pharmacology, College of Medicine, University of Arizona, Tucson, Arizona 85724-5050, USA.

Tales from a continental drifter

The Eternal Darkness

by Robert Ballard with Will Hively
Princeton University Press: 2000. 388 pp.
\$29.95, £18.95

Alan Longhurst

Matthew Maury, the founding father of oceanography, knew as early as 1850 that a ridge on the sea floor ran the length of the Atlantic Ocean, but it would be another century before its real significance was generally accepted. In his 1882 novel *Atlantis: The Antediluvian World*, Ignatius Donnelly identified the new ridge with Atlantis, Plato's lost,

rich island submerged by an 'earthquake'. Donnelly invoked antique literature, anthropology, linguistics, folklore, artefacts and architecture to propose an Atlantean cultural cradle for both Europe and Meso-America. He was taken seriously in his time, for my copy is the book's 24th edition.

Of course, earlier writers had already understood that the shape and arrangement of the continents could not be accidental and had pondered the reason. But even after Frank Taylor, Howard Baker and Alfred Wegener had independently and almost simultaneously (1908–12) proposed that continental masses must have shifted their relative positions, geologists remained deeply divided. Indeed, as late as 1937, Alex du Toit's crystal-clear book *Our Wandering Continents* (Oliver & Boyd), under the motto "Africa holds the key", still failed to move some.

It was only in the 1960s that fast-accumulating evidence caused the nineteenth-century doctrinaire castle — built on Sir Charles Lyell's *Principles of Geology* (1830–33) — to crumble. How Marie Tharp's compilations of all the available bathymetric soundings of the Atlantic Ocean's floor revealed the rift valley, and how Bruce Heezen was converted to the continental-drift theory after considering the East African rift, is an often-told story.

Which brings us to Robert Ballard's book, which tells the story all over again. Ballard is a slightly larger-than-life US Navy Reserve officer, geologist and discoverer of *Titanic*, who knows as well as anyone how to move and shake the system to get what he wants. And what he has always wanted is a deep-diving submersible.

Like those who shook the Royal Society and the Admiralty and sent HMS *Challenger* on its long voyage in 1872, Ballard has successfully persuaded the authorities — in his case, the US Navy — to support his deep-ocean research since the early 1970s. He had impeccable timing, because around then oceanographers in the United States were hit by a tidal wave of funding from the Office of Naval Research and the National Science Foundation. It was a giddy time that will not soon recur. But, because modern navies lose things on the deep-sea floor and want to find them again, Ballard has had no serious difficulty in subsequently keeping his hands on a deep-diving submersible.

As the 1970s were also the years when the significance of mid-ocean ridges was finally realized, Ballard was able to fund their direct exploration. He was one of the fortunate few who visited rift valleys during the US and French expeditions using the submersibles *Alvin* and *Cyana*, and was present at the first observations of the spectacular fauna at the hot vents, for which names such as "Rose Garden" and "Garden of Eden" were inevitably coined. In their highly manoeuv-

vable submersibles, the biologists were able — as Jacques Cousteau had said that they must be one day — to "fly like angels" around this newly discovered paradise. Although it smelt of rotten eggs when the water-sampling flasks were opened, this paradise was the energy source for both free-living and symbiotic sulphide-oxidizing bacteria.

Other popular books have done a better job on the scientific consequences of the Tharp–Heezen revelation, and I much prefer the description of both plate tectonics and the vent fauna in Robert Kunzig's *The Restless Sea* (Norton, 1999). As in most popular accounts, Ballard overemphasizes the metabolic isolation of the vent fauna from solar energy in favour of energy from geothermal sources. Of course, free oxygen is required to release the energy of the sulphide bond, and this must originate from photosynthesis in green plants.

Naturally, Ballard's account dwells on the submersibles that were used in these explorations. He introduces both the bathysphere used by William Beebe and Otis Barton in the 1930s, suspended on a very long wire, and the poorly manoeuvrable bathyscaphes, buoyed by large tanks of petrol, that were pioneered by Auguste Piccard and his son Jacques in the 1950s. The first recognizably modern vehicle was Jacques Cousteau's shallow-diving and long-lived 'diving saucer' of 1959, which was widely copied in the 1970s when aerospace briefly got into the deep-diving act. Of these later submersibles, only a few survived the competition for funding, notably the workhorses *Alvin* of Woods Hole Oceanographic Association in Massachusetts and *Cyana* of the French Centre National de la Recherche Scientifique. Their exploits on the mid-ocean ridges and elsewhere are chattily recounted.

Recently, a trend to replace manned submersibles with remotely controlled devices, such as *Argus* and *Jason/Medea*, has been driven by developments in digital electronics and optical-fibre communication. Ballard is a strong advocate of remote control and observation as the way of the future — in which he is probably correct — but others remain faithful to manned vehicles. Whether remote observation leading to mass participation in distant auditoria is anything more than public-relations fluff, only time will tell, but it is available if needed for anything serious.

For the US Navy, Ballard developed search techniques for locating things on the deep-sea floor, including the reactor of USS *Thresher*, which sank in 1963 off the coast of New England. This work led him eventually and enthusiastically to *Titanic*. The book's back-cover blurb calls this "one of the most significant discoveries of all time", but I call it voyeurism and I didn't enjoy the whistle-blowing, flag-waving return of the expedition to Woods Hole after the news of their finding

Titanic had created a media storm. A videotape of the wreck, handed up to an opportunist long-range helicopter and addressed rather pompously "To the World Press", had been quickly broadcast as a "CBS exclusive" — a silly business that predictably upset Ballard's French partners in the expedition.

Knowing some of his scientific papers, I judge that there is not enough Ballard (or too much Hively) in this book. There is also some thoughtless editing, especially in the figures — some need scales, or are over-reduced or have unhelpful legends, while a few are downright misleading. Oddly, there are 60 pages of "further reading", simply listing references to the primary literature, although not one is cited in the text. Perhaps they are the contents of Ballard's reprint boxes, given to his ghost-writer for background, and arranged sequentially by year? ■

Alan Longhurst is at Place de l'Eglise, 46160 Cajarc, France.

Winging their way

The Biomechanics of Insect Flight: Form, Function, Evolution

by Robert Dudley

Princeton University Press: 2000. 476 pp.

\$49.50, £31

Form and Function of Insect Wings: The Evolution of Biological Structures

by Dmitry L. Grodnitsky

Johns Hopkins University Press: 1999. 261 pp.

\$49.95, £41.50

R. McNeill Alexander

Most animal species are insects, and the most conspicuous characteristic of the insects is that they fly. Research on insect flight has been held back by approaches that, with hindsight, seem crude and inappropriate, but the past two decades have brought many exciting advances.

Previously, we thought of insect wings as stiff, flat plates — now we know that some bend and twist in flight in ways that must have large aerodynamic effects. It used to seem prohibitively difficult to make physiological measurements on free-flying insects, or even to obtain film of adequate quality for detailed analysis of the movements of free flight, so most observations were made on insects induced to flap their wings while glued to supporting wires. Now we can measure the oxygen consumption of a single bee flying free in a wind tunnel, and even more ambitious experiments are in prospect.

Another approach now seen to have limited value is the use of blade-element theory, which works for helicopter rotors, to estimate aerodynamic forces on insect wings. Calculations give the maximum forces the wing motion should be able to generate. In

many cases, these are much less than is needed to keep the animal airborne, inspiring the dictum that engineers have proved bumblebees cannot fly. Bumblebees have, of course, convincingly proved otherwise.

It has long been realized that the crucial difference is that aircraft aerofoils move more or less steadily through the air, whereas the flapping wings of insects repeatedly reverse their direction of motion, giving rise to effects that are not predicted by 'steady-state' aerodynamics. Blade-element theory can do no more than tell us when steady-state aerodynamics is inadequate; even when it shows that steady-state forces could support the insect, it cannot tell us that no unsteady effects are acting, and it can never

tell us what the unsteady effects are. The focus has recently shifted from the movements of the wings to the movements of the air they drive, leading to qualitative explanations of aerodynamic mechanisms that will soon, one hopes, become quantitative.

Much less progress has been made in the possible energy-saving role of elastic structures. It has long been known that some insects could save a lot of energy if the kinetic energy taken from the wings at the end of each stroke could be stored and returned in an elastic recoil. Arguments that this does or does not happen have generally depended on estimates of the efficiency with which the flight muscles must work. If the assumption of no elastic storage led to implausibly high



Flight recordings

One of photographer Stephen Dalton's first assignments, in the 1960s, was a book about honeybees. His discovery that insect flight had never been clearly caught in a still photo launched both an obsession with solving this problem and an award-winning career. The results, building on recent advances in high-speed photography, are a highlight of his latest

book *Secret Worlds* (Firefly, \$35). In 125 colour photos, each with explanatory text, some animals are seen hiding or resting. But Dalton's fascination with capturing movement creates the most memorable images: the agility of a diving vole, a cockchafer coming in to land, the aerial manoeuvres of bats, and the astonishing sight of a basilisk running across water.

book reviews

estimates of efficiency, it was concluded that elastic mechanisms must operate. Unfortunately, it has not been clear how high an efficiency is plausible, and there have been no satisfactory calculations of energy savings based on measurements of the elastic properties of the thorax. An inspired new technique is needed to advance our knowledge, but 'work loop' experiments have made important steps in the right direction. In these muscles are stimulated while being stretched and allowed to shorten at rates that simulate the action of flight.

Robert Dudley has written a remarkably comprehensive account of our knowledge of insect flight: his bibliography fills 102 pages. No other recent book covers the field so fully, so this one will be most valuable to students and researchers. But his aim seems to have been to inform rather than to teach. The book reads like a series of review articles on the morphology, evolution, aerodynamics, physiology and control of flight. There are remarkably few illustrations and too little mathematics. There is frustratingly little detail about the methods and results of even the most significant experiments. However, he has tried to help by providing chapter summaries, a list of symbols and a glossary.

Dmitry Grodnitsky has written a book of much narrower scope, with far more experimental and morphological detail. It starts with the experiments in which he and his colleagues visualized the air movements in the wake of tethered insects beating their wings in still air. Observations on free flight would have been preferable, had they been feasible, but these demonstrations of vortex structure are nevertheless valuable. The book continues with discussions of the evolution of insect flight, of the functional morphology of the veins and folds of insect wings, and of the scales on the wings of butterflies and moths. This is a book for specialists. ■

R. McNeill Alexander is in the School of Biology, University of Leeds, Leeds LS2 9JT, UK.

Strolling through the bioenergy field

The Energy of Life

by Guy Brown

Free Press: 2000. 288 pp. \$25, £16.99

Paul D. Boyer

Guy Brown defines energy loosely. "We have many words to express the high energy state: vitality, vigor, vivacity, strength, arousal, ardor, drive, fervor, stamina, gumption, zeal and zest," he states. Such a definition allows the author to explore the basis of many aspects of human behaviour in this unusual and readable book.

Brown is a respected researcher in the

field of bioenergetics and is well served by his broad knowledge and historical perspective. His introductory descriptions of how living processes are made possible by the wondrous machinery of the cell are what I would like to give my non-scientific friends who ask what my work is about. Avoiding technical details, he describes such processes as how DNA is a blueprint for some 50,000 proteins that make possible the occurrence, control and complexity of metabolic pathways, and that energy is captured in the structure of a molecule called ATP.

In a chapter on "The body electric", however, he appears over-enthusiastic in the use of the term 'electricity', including not only the transfer of electrons, but also the use of proton and sodium gradients as proton and sodium electricity and the use of ATP as phosphate electricity. Most bioenergeticists would feel the statement "Our cells are energized by huge electric fields driving currents of charged particles via a myriad minuscule wires" does not convey a proper perspective.

Brown is writing for a general audience, though even graduate students in biochemistry would enjoy his historical accounts of the development of current understanding in such areas as glycolysis, the electron-transport chain, the capture of energy by development of electrochemical proton gradients, and the use of phosphate-bond energy. Other nuggets scattered throughout the book will interest and entertain readers in the fields of bioenergetics and biochemistry.

Mitochondria, the cellular organelles in which ATP is made, were introduced into our cells billions of years ago by fusion with a microorganism. It is from these organelles that free radicals escape and damage our DNA and proteins, eventually leading to ageing. As Brown emphasizes, a wandering lone electron may wreak havoc with hundreds of molecules. Although antioxidants do play a preventive role, he inappropriately implies that purified vitamins C and E and beta-carotene are less effective than natural antioxidants. It is the structure, not a 'natural' source, that is important. The tendency of mitochondria to leak protons and radicals is regarded as a fault in nature's design. But another view is that they are a splendidly honed organelle that captures most of the energy released by use of toxic oxygen, with enough preventive and repair mechanisms to allow us to reach old age.

Brown ranges widely, from motion to metabolic rate and on to obesity, including chapters on mind energy, brain waves, and sex and sleep. He recognizes that social, psychological and neurophysiological factors all play a role in the pace of life. (I learned that giving rats a shot of adrenaline right after a learning experience increases memory retention.) We can all relate his good discussion of the effect of stress on metabolism and behaviour to our own experiences.

We learn that Sigmund Freud was not a well man (he suffered from chronic fatigue and nervousness, and was enthusiastic about cocaine); that the term 'mesmerism' was derived from Viennese physician Franz Mesmer, who attributed powers to a force called animal magnetism; that Benjamin Franklin headed a commission which concluded that such a force did not exist; and that for better sex one should try the daytime or early evening. In the chapter on sex and sleep, Brown even discusses life in the scientific community. Discoveries such as that of nitric oxide as a signal transmitter create a "wave of interest in the scientific community"; if you succeed in getting on to the wave of discovery, your career starts to flourish.

A 58-page appendix, "The story of living energy", does not meet the objective stated in the introduction as giving the background ideas in bioenergetics. Instead, it is an elegant historical account of how an understanding of life's processes has been acquired, from the insights of ancient Greece to the recognition that fermentation is a function of living cells. Readers are exposed to such novelties as the type of healthy regime Socrates might have prescribed.

The very wide range of topics Brown considers, and his slight tendency to wander and repeat himself, do not detract importantly from this informative and enjoyable account of the scientific basis of our everyday life and behaviour. ■

Paul D. Boyer is in the Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, USA.

New in paperback

The Unnatural Nature of Science

by Lewis Wolpert

Faber & Faber, £8.99

Living with our Genes: Why They Matter More Than You Think

by Dean Hamer & Peter Copeland

Pan, £6.99

Hermaphrodites and the Medical Invention of Sex

by Alice Domurat Dreger

Harvard University Press, £10.50, \$16.95

"Ignore the somewhat trendy title, which might seem to threaten yet another piece of pretentious postmodernism. This is a well-researched, sober history of a problem that Alice Dreger shows has directly affected more people than we might think and which shapes the sense of sexual identity of us all ... Dreger shows how deeply ingrained are our assumptions about gender normality (sexual anatomy is destiny), and on how flimsy a basis they have been grounded. The book offers us all a lesson in self-awareness." Roy Porter, *Nature* 393, 323 (1998)

Oeconomia Naturae L.

The ecology of Linnaeus was Stoic, Baroque and surprisingly modern.

Geir Hestmark

The division of labour was clearer in the eighteenth century: the supervisor wrote the thesis, the student defended it in an academic disputation. (Some supervisors may feel that it still works this way.) And so it came to pass that on 4 March 1749, in the idyllic Swedish university town of Uppsala, the 24-year-old Isac Isacsson Biberg at the Faculty of Medicine defended a *Specimen academicum de Oeconomia Naturae* under the Praesidio of Carl Linnaeus.

For Linnaeus it was becoming routine. Between 1743 and 1776 he wrote more than 180 such academic theses. But few achieved the instant success of the *Oeconomia Naturae*. A Swedish translation was produced within a year. English and German versions soon appeared. It was also reprinted in Latin in the many editions of Linnaeus's *Amoenitates academicae* published in Amsterdam, Leyden, Erlangen and Graz through the second half of the eighteenth century. New translations continue to appear today.

Oeconomia Naturae is both the culmination of a great tradition — that of Christian natural theology, and the starting point of a new science, the one that Ernst Haeckel named 'ecology' in 1866. In accordance with

the natural theology and the 'age of optimism' celebrated in the works of William Derham, John Ray, Bernhard Nieuwentyt, Gottfried von Leibniz and Christian von Wolff, Linnaeus defines 'the economy of nature' as the Creator's wise arrangement and deposition of all things according to which they fulfil their purpose for the glory of God and the happiness of Man.

The great wave of optimism, however, was already crumbling around 1750, and the nature of Linnaeus seems ruled more by fate and destiny than providence and luck. This emphasis (which Linnaeus would pursue to the bitter end in his notebooks on *Nemesis Divina*) reflects the powerful influence that ancient Stoicism, notably the works of Seneca, had on him. But it is also very much the view of a faithful observer of nature's ways. Linnaeus's acute and comprehensive first-hand experience sets him apart from the previous physico-theologists, who were more philosophers than naturalists.

Stoicism and Christian natural theology do, however, share a common spiritual aim: peace of mind, achieved by acceptance rather than despair, resignation rather than revolt. Both systems attempt to transform and transcend the pain and misery of life on Earth — the 'Nature red in tooth and claw' that Alfred

Lord Tennyson would lament in his *In Memoriam* 100 years after the *Oeconomia*.

Taking as its point of departure a motto from Seneca, "Everything is forever changing", the *Oeconomia Naturae* revolves around the fate of the individual, the fundamental processes or stages in the life cycle: creation/birth (*Generatio*), maintenance (*Conservatio*) and finally destruction (*Destructio*) — in plants, animals and rocks. Linnaeus paints a vivid picture of how these processes create a flow of matter through nature — what we would refer to as the great biogeochemical cycles — so that everything is connected and nothing is really lost.

And although individuals perish, their roles persist. Linnaeus was a child of the late Baroque, and his *Oeconomia Naturae* is an exposition of what is often called the fundamental metaphor of that age, the 'Theatrum Mundi', encapsulated by the Spanish Baroque poet Calderón de la Barca's play *El Gran Teatro del Mundo* (The Great Theatre of the World). The world is a stage, where, for a brief, almost dreamlike moment, each individual partakes in the great play of being. This is a recurrent metaphor in Linnaeus's work — used, for instance, in the introduction to the tenth edition of his *Systema Naturae* (1758), the starting point of zoological nomenclature.

The roles in Linnaean nature are what today's ecologists call 'niches': a multidimensional 'space' defined by the abilities of the species and their interactions with the environment — their physiology and habitat preferences, position in food chains and ecosystem structure. Although the *Oeconomia Naturae* reads like an ecology textbook, it also sparkles with the eroticism of the Baroque. Like a voluptuous painter, Linnaeus revels in the splendour of life, in its beautiful 'costumes', its sensual appeal and showy extravagance, the delightful colours, forms and adaptations, the impressive devices for preservation, survival, defence, attack, sex and propagation, mating and pollination, the means of dispersal and child-rearing.

At the turn of the millennium, both Stoicism and a Baroque voluptuousness (consumerism) are back in vogue in western culture. Yet the greater economy of nature Linnaeus praised is being depauperated by the catastrophic rate of species extinction. The roles are getting fewer, the Theatrum Mundi and life on Earth impoverished. To stop this destruction will be a major — perhaps the major — challenge of the new millennium. ■

Geir Hestmark is in the Department of Biology, University of Oslo, 0316 Oslo, Norway.



Primary producer: Linnaeus — here surrounded by Baroque angels — wrote about 200 academic theses, combining natural theology with keen observation of nature. Illustration from Robert John Thornton's *Temple of Flora*.



Et in articulo mortis

Humankind still has need of dragon slayers.

Han Yu

The Chinese had a tradition about eclipses. Such things happened, they said, when a dragon tries to swallow the Sun: but the dragon would go away if people shouted at it. The Chinese were, of course, correct. There really are dragons that swallow suns, and they will go away — but you have to shout very loudly indeed.

The first signs of dragons came in 2618 of the Common Era (Nebular Year 4667532109 $\pm$ 1044 Calibrated), when several nearby stars ceased to exist. The first were red dwarfs far beyond naked-eye visibility.

A greater disturbance came in 2633 with the extinction of Lac 9352, a ninth-magnitude class-M2 dwarf less than 12 light years from the Sun. This star had three Earth-like planets that the Yerkes Automatic Heliospheric Optical Observatory (YAHOO) was mapping, when the star simply vanished. Instrument malfunction was blamed at first, but analysis of YAHOO images taken just before the event showed several black shapes converging on the star. And then, in 2634, Sirius vanished.

After that, things happened very quickly. After centuries of quarrel, the various schisms of humanity agreed to move the entire population of the Solar System into spun-geode asteroid habitats. In 2667, humanity started to quit the home of its birth.

Only after the complete depopulation of the System, and the observation (in the rear-view mirror, as it were) of packs of dragons approaching the Sun — were other seemingly unrelated facts considered. The Exodus Fleet was a vestige of a species reduced by a disease which was only much later connected to the plague of dragons.

The first case of postembryonic oolithic petrosis (POP) was recorded in a military hospital on Phobos in 2596 during the Sec-

ond Palladian War — humanity's final internecine conflict. POP was, at first, thought to have been a contagion bred in the death camps of the late Director of Pallas, Napoleon Ireneo Funes III.

The first victim of the disease stiffened, as if frozen. His skin became covered in scaly welts that grew until they became confluent, swaddling him in a jet-black shell of adamantine hardness. At first the shell measured the contours of the body beneath, but then contracted to become a perfect, utterly inert, X-ray-opaque sphere 35 centimetres across. The corpse (if that is what it was) was dissected — but the contents were found to be a gelatinous matrix secreted by roving amoebocytes.

Other cases of POP came to light around 2600, at first in ones and twos, and then in whole communities. Despite intense work, no cause or trigger was identified. After about 30 years, POP took off in a terrifying, surging pandemic. By 2632, the populations of the Moon, Iapetus, Mercury, Callisto and Australia ceased to respond to communication. A Mesoamerican expedition to Britain in 2635 reported the country deserted, the population reduced to quiescent, black spheres.

By 2667, the population of the Solar System had shrunk to the two billion of the Exodus Fleet — the remainder left behind, the unresponsive globes thought to represent the terminal phase of POP. The last habitat in the Exodus Fleet crossed the heliopause in 2699.

Finally left to themselves, the spheres — the remnants of nine billion souls — changed. The gelatinous contents reorganized and developed. Genes in the innocuous-looking amoebocytes — genes silent for the entire span of organismal evolution and thought to be 'junk' — were, at last, transcribed. Within the core of each sphere, matter itself changed its shape, unlocking tiny doors into the heart of the cosmos.

In 2815, the Alpha Centauri system was consumed by a pack of at least thirty thousand dragons. When the signal reached the Solar System in 2819, there were no eyes to see. But it did not go unnoticed. As if on command, nine billion spheres cracked open. As one, they rose into space, each surrounded by an actinic aura and spitting gamma radiation. As if choreographed, nine billion spheres converged in space to form a cordon around the Sun.

In 2845, dragons entered the Solar System and joined battle. The dragons were defeated, but at great cost. The effect on the Sun of absorbing 30,000 dragon carcasses, each one an organized lattice of neutronium, was predictably catastrophic. The Exodus Fleet was hurried on its way by a supernova blast wave. Even had it wanted to, humanity had no home to come back to.

Years later, we observers on the Exodus Fleet finally put the facts together. The key was found in ecology. Over decades, ecologists had amassed cases of prey species which, when threatened by even a hint of an approaching predator, would change dramatically in form and turn against the aggressor.

The first brood of dragons hatched when the Universe was born. The living creatures on the early generations of planets were easy prey, but natural selection worked its remorseless logic. Planets of species evolved to fight dragons by generating castes of dragon slayers, leaving a small inoculum to spread the species to other planets, other stars.

We were just such an inoculum, furthering life in the cosmos, keeping one step ahead of the dragons. POP was not a disease. It was part of the natural order of the Cosmos. And so will it be until Ragnarök.

Han Yu lives in Cupertino, California. A collection of essays, *The Ghost of Jorge Luis Borges Which Can Also Be Used As a Table*, is published by Unicorn Gardens Press.

Evolution all at sea

Richard D. Norris and Colomban de Vargas

Evidence of gene flow between plankton in the Arctic and Antarctic is one surprise. Another is the discovery of hitherto hidden genetic diversity in these organisms.

Planktonic organisms evolve in continuous motion. Drifting passively in oceanic currents, their distributions seem to be limited mostly by changes in the temperature, nutrients and structure of the water masses in which they live. Or are they? The long distances separating species in the Arctic and Antarctic, and the seemingly substantial barrier imposed by the warm tropical oceans, offer a key test of the frequency of long-distance dispersal in the oceans and its importance in evolution.

Planktonic foraminifera are microscopic organisms found throughout the world's oceans. Some species are bipolar — that is, they occur at opposite ends of the Earth in the chilly waters of the Arctic and Antarctic. On page 43 of this issue, Darling *et al.*¹ show for the first time that particular variants of bipolar species have such striking genetic similarity that they must regularly exchange populations through the tropical seas. At the same time, the data reveal considerably more genetic diversity in these polar species than was ever expected from studies of their skeletons. The new results constitute a powerful argument that the open oceans contain more species, with much better dispersal capabilities, than previously suspected.

The discovery of so much hidden biodiversity in the foraminifera is not merely of academic interest. There is a continuous rain of the calcareous shells of these globally distributed, short-lived and fecund organisms from the surface waters to the ocean floor. The result is a superb archive of climate history over the past 130 million years²: analyses of the abundance and geochemistry of foraminiferan shells are standard tools for reconstruction of ocean circulation and climate. But if a 'species' is actually three or four different species, each with a unique ecology and environmental distribution, then we have been overlooking a vast storehouse of information about the ancient oceans. Indeed, by learning to recognize the different genetic varieties of planktonic foraminifera, it may be possible to improve dramatically the resolution of climate proxies.

Darling *et al.*¹ analysed genetic relationships between three high-latitude 'species': *Neogloboquadrina pachyderma*, *Globigerina bulloides* and *Turborotalita quinqueloba* (Fig. 1). These are among the most abundant of high-latitude foraminifera and are key indicators of temperatures, nutrients and

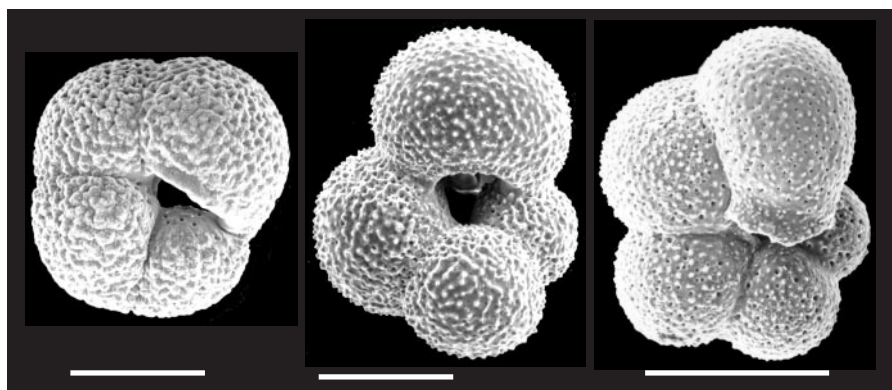


Figure 1 The three species of foraminifera studied by Darling *et al.*¹. Left to right: *Neogloboquadrina pachyderma*, *Globigerina bulloides* and *Turborotalita quinqueloba*. Darling and colleagues' data¹ suggest that each of these 'species' can be subdivided into three to five distinctive genetic variants, of which several are found in both the Arctic and the Antarctic. Scale bar is 100 μm .

salinity in subpolar to polar surface waters. Comparisons of live specimens from the Arctic and Antarctic show that each 'species' consists of three to five distinctive genetic varieties. *N. pachyderma* exists as left-coiling and right-coiling forms, and the marked genetic differences between them belie the common assumption that the different coiling results from growth of populations of the same species at different water temperature. Likewise, high-latitude varieties of *G. bulloides* and *T. quinqueloba* are distinctly different from populations in the subtropical ocean.

Some of the other genetic variants in each 'species' come only from the Arctic or the Antarctic. Nonetheless, and most strikingly, one or two are present at both poles. Complete DNA sequence identity between specimens from both poles implies that there has been genetic exchange between Arctic and Antarctic populations since the establishment of the modern polar provinces 16–18 million years ago. However, estimation of the time when gene flow occurred between the bipolar populations is constrained by the slow rates of evolution in the gene fragment studied by Darling *et al.*¹. In our view, the data suggest there has been genetic exchange between the poles some time in the past 200,000 years. Indeed, it is possible that exchange is occurring today but it will take studies of more rapidly evolving parts of the genome to find out for sure.

Just how polar species cross the tropics is unclear. But there are clues. Today, upwelling of cold, nutrient-rich waters occurs

seasonally along the western margins of the Americas and Africa. Polar species could use upwelling cells as stepping stones to move across the Equator. Genetic exchange between the poles may also have been easier during glacial stages when polar waters shifted towards the Equator and upwelling areas in the tropics expanded.

Because barriers such as the warm tropics do not seem greatly to restrict genetic exchange between the poles, how do new species arise in the open ocean? Oceanographic barriers to dispersal have long been assumed to restrict gene flow between populations, and create genetic isolation and divergence of new species. But speciation may rely more on the erection of biological barriers to reproduction than on the prevention of population exchange³. Populations could be genetically isolated from one another by synchronizing reproduction⁴ to different climatic or ecological signals. Incipient species may also differ in the biochemical cues that allow their gametes to recognize one another⁵.

There is increasing evidence that many genetic varieties of planktonic foraminifera are ecologically distinctive and may represent different species^{6–8}. For example, we have found three highly divergent genetic types of *Orbulina universa*, each of which is associated with waters of different productivity⁷ (Fig. 2, overleaf). Likewise, four slightly different genetic groups have been detected in the relatively young species *Globorotalia truncatulinoides*, which successively colonized different southern water

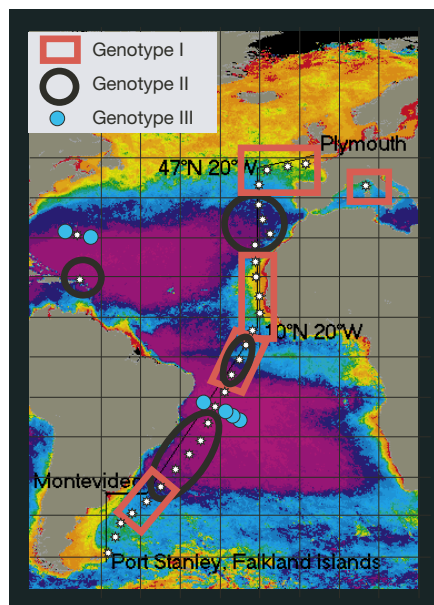


Figure 2 Ecological differentiation in three distinct genetic groups of the subtropical-tropical planktonic foraminifer *Orbulina universa*⁷. One variant (genotype I) was collected from areas with high chlorophyll concentrations, indicating high biological productivity (red boxes); the second (genotype II) came from waters with lower chlorophyll concentrations (black circles and ovals); and the third (genotype III) was found only in low-productivity waters of the subtropics (blue dots).

masses⁹. There are hints that the high-latitude species described by Darling *et al.*¹ also have different ecological preferences. Apparently, the various genetic types have become specialized in different surface-water masses from the tropical to the polar oceans.

Clearly, in traditional morphological analyses of planktonic foraminifera, a great deal of diversity has been overlooked or regarded as mere variation. It appears that each of the 50 or so living species² is likely to represent clusters of related species. One need will be to find ways of recognizing the various genetic types in the fossil record so that they can be used to reconstruct ancient environments and global climate more accurately. Indeed, the varieties of *Orbulina* appear to be recognizable by their shell porosity, as is the case for other genetically distinctive variants of foraminifera⁸. Limited genetic studies of different groups, such as the phytoplankter *Prochlorococcus*¹⁰ and the fish *Cyclothone*¹¹, suggest that the underestimation of marine diversity is widespread. Evolutionary models that rely on barriers to dispersal in the oceans may well need a fundamental rethink in light of these new data.

Richard D. Norris is at Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543-1541, USA.

e-mail: rnorris@whoi.edu

Colomban de Vargas is in the Département de

Zoologie et Biologie Animale, Université de Genève, Station de Zoologie, 154 route de Malagnou, 1224 Chêne-Bougeries, Genève, Switzerland.

e-mail: colomban.devargas@zoo.unige.ch

1. Darling, K. F. *et al.* *Nature* **405**, 43–47 (2000).
2. Hemleben, C., Spindler, M. & Anderson, O. R. *Modern Planktonic Foraminifera* (Springer, New York, 1989).
3. Norris, R. D. in *Reconstructing Ocean History: A Window into the Future* (eds Abrantes, F. & Mix, A.) 173–193 (Plenum, London, 1999).

4. Palumbi, S. R. *Annu. Rev. Ecol. Syst.* **25**, 547–572 (1994).
5. Biermann, C. H. *Mol. Biol. Evol.* **15**, 1761–1771 (1998).
6. Darling, K. F., Wade, C. M., Kroon, D., Leigh Brown, A. J. & Bijma, J. *Paleoceanography* **14**, 3–12 (1999).
7. de Vargas, C., Norris, R., Zaninetti, L., Gibb, S. W. & Pawlowski, J. *Proc. Natl Acad. Sci. USA* **96**, 2864–2868 (1999).
8. Huber, B. T., Bijma, J. & Darling, K. *Paleobiology* **23**, 33–62 (1997).
9. de Vargas, C., Renaud, S., Hilbrecht, H. & Pawlowski, J. *Paleobiology* (in the press).
10. Ferris, M. J. & Palenik, B. *Nature* **396**, 226–228 (1998).
11. Miya, M. & Nishida, M. *Nature* **389**, 803–804 (1997).

Palaeoanthropology

Coasting out of Africa

Chris Stringer

It is now generally accepted that Africa is the ancestral homeland of modern humans, *Homo sapiens*^{1–3}. But the timing of human dispersal from Africa, and the routes taken, remain controversial. Was there only one major dispersal, which took place after 50,000 years ago¹? Or were there several episodes of migration, starting earlier and covering a longer period of time²? Moreover, was the most obvious route taken — through Sinai and the Levant (the eastern Mediterranean coast) — or might there have been other pathways?

On page 65 of this issue, Walter *et al.*⁴ present the first well-dated evidence that humans were living along the African coast of the Red Sea during the last interglacial, around 125,000 years ago, and were probably exploiting marine food resources. They argue this finding implies that there was a major change in human adaptive capacities at about that time. It may also indicate that coastal routes were used by early *Homo*

sapiens to leave the African homeland. Other shorelines of Late Pleistocene age (oxygen isotope stages 2–5; see Fig. 1 for chronology) might therefore preserve further traces of our exodus from the continent.

At one time it was believed that humans began to exploit marine resources only during the Eurasian Upper Palaeolithic, after 40,000 years ago⁵. But it seems that both early modern humans in southern Africa and Neanderthals in the Mediterranean did this during the Middle Palaeolithic^{1,6}, at sites such as Klasies and Herold's Bay Caves in South Africa, and Vanguard Cave (Gibraltar) and Moscerini Cave (Italy), respectively (Fig. 2). However, much of the Late Pleistocene evidence of these seaside dwellers must now be submerged beneath the high sea levels that have pertained during the present interglacial, which started some 12,000 years ago. Because of their elevated altitude, due to the higher sea levels of that time, sites of last-interglacial (early oxygen isotope stage 5) age

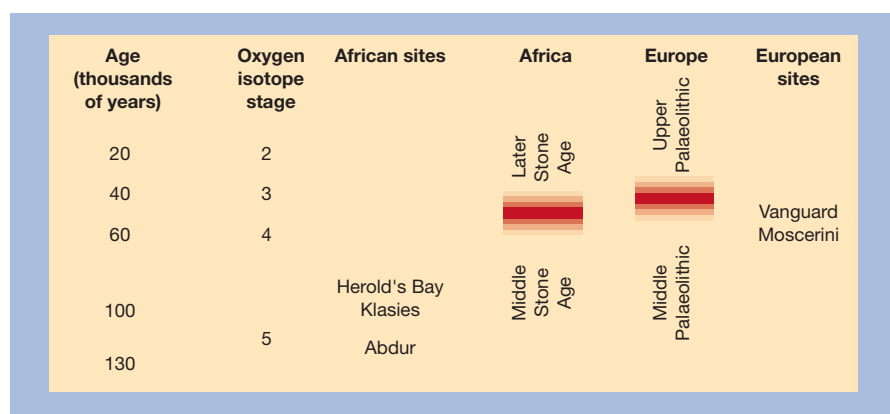


Figure 1 Chronology for the Late Pleistocene¹. Oxygen isotope stages reflect climatic changes inferred from isotope ratios in marine microorganisms in deep-sea cores (roughly, odd numbers indicate relatively warm conditions and even numbers relatively cold). The sites mentioned, including Abdur⁴ (discussed here), are those with evidence of the early exploitation of marine food resources. The traditional division of the European Old Stone Age is into Lower, Middle and Upper Palaeolithic stages. In Africa, a comparable scheme uses the terms Early Stone Age, Middle Stone Age and Later Stone Age, and there is some evidence that the African transition between the Middle and Later Stone Ages occurred before the European transition from Middle to Upper Palaeolithic. Middle Palaeolithic industries in Europe were produced by Neanderthals, whereas Middle Stone Age industries were probably the product of modern humans.

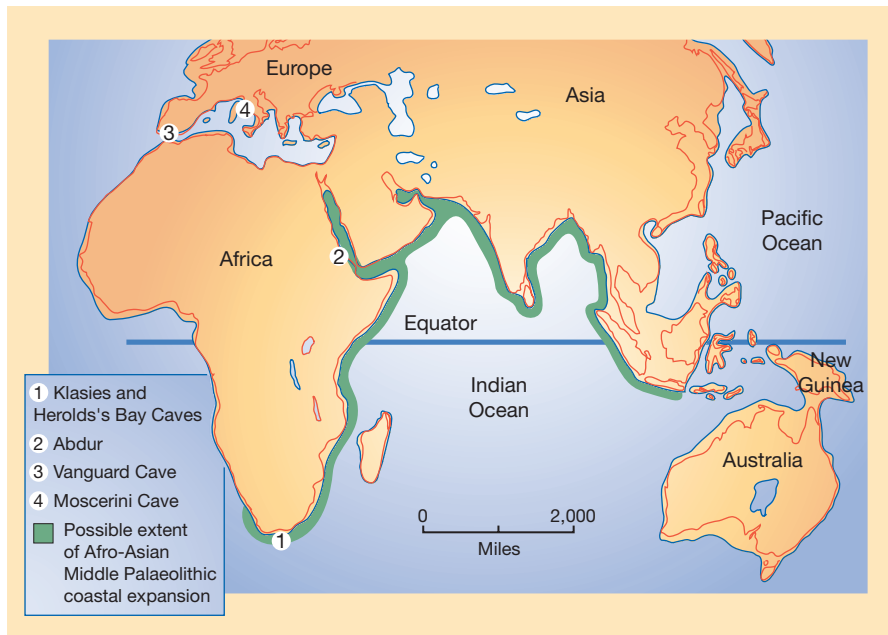


Figure 2 The Old World as it might have looked during the Late Pleistocene, around 65,000 years ago. At that time, large amounts of sea water were locked up in polar ice sheets, so sea levels were lower and more land was exposed (present-day shorelines are in red). Known coastal sites of human occupation during the Middle Palaeolithic are marked; Abdur is the site investigated by Walter *et al.*⁴ and discussed here. A putative route for coastal migration of modern humans from Africa to Asia is shown in green. (Redrawn after ref. 10.)

are more likely to have remained exposed. This is particularly true where the land has been uplifted by geological processes, as is the case around parts of the Red Sea.

Walter *et al.*⁴ recovered Middle Stone Age (African Middle Palaeolithic) artefacts, such as hand axes and obsidian flakes, from strata in a raised fossil reef near Abdur in Eritrea (Fig. 3). Geomorphological considerations and correlation with other Red Sea localities suggest that the site dates to the last interglacial. Walter and colleagues confirmed this age with uranium-series dates that, on average, gave a figure of about 125,000 years. Who made the artefacts is unknown. But there are fossils of near-modern or modern *H. sapiens* from around this time^{1,2} in neighbouring regions such as Ethiopia (Omo Kibish), Sudan (Singa), Kenya (Guomde) and Israel (Skhul and Qafzeh). So it is likely that the people concerned were early members of our species.

Klein¹ has argued that the main dispersal of modern humans from Africa probably occurred only after the beginning of the Later Stone Age (equivalent to the Eurasian Upper Palaeolithic). For him, that event heralds the beginning of modern cognitive and adaptive capabilities. According to this view, then, the presence of modern humans in the Levant during the last interglacial, represented by the burials at Skhul and Qafzeh, was only a brief geographical extension of the species from Africa. The real dispersal of *Homo sapiens* was through that region, but did not occur until the Upper Palaeolithic, perhaps 45,000 years ago.

Other workers have favoured an earlier, Middle Palaeolithic, beginning for dispersals. Kingdon⁷ proposed that Middle Palaeolithic people left Africa through the Levant and reached southeast Asia by 90,000 years ago. There they adapted to coastal conditions, and developed a boat- or raft-building ability that enabled them both to return to Africa and to move southwards to Australia. By contrast, Lahr and Foley² suggest in their 'multiple dispersals model' that a more direct route from Africa to Arabia and further east could have been taken before 50,000 years ago, perhaps using the coast. However, subsequent dispersals to the north, evidence for which comes from early Upper Palaeolithic artefacts found in countries such as Egypt, Israel and Bulgaria, would have followed the Levantine route.

The findings of Walter *et al.*⁴, together with new data from Australia, allow further elaboration of these possibilities. There is increasing archaeological evidence⁸ that Australia was colonized (by boat, because no landbridges existed during the Pleistocene) before 50,000 years ago — that is, before the proliferation of Later Stone Age and Upper Palaeolithic features such as blade tools and art. Moreover, a modern-human burial site from southeastern Australia, associated with the symbolic use of red ochre, has been re-dated to about 60,000 years ago⁹. This implies that at least one dispersal of modern humans from Africa must have occurred during the Middle Palaeolithic, and that characteristic



100 YEARS AGO

The motion for the second reading of the Sea Fisheries Bill in the House of Commons, on Monday, resulted in a lively discussion. The Bill prohibits the sale of flatfish below a specified size, and its rejection was moved on the grounds that it would not have the effect of preventing the destruction of immature fish, or of increasing the supply of fish. In the course of discussion, an honourable member said that the whole of the trouble arose from the institution of a number of committees composed of farmers, lawyers and captains of the horse, foot, and artillery, who knew little of fishing, and who ventilated strange theories and supported them with portentous and irrelevant statistics. This remark was used as an argument against the Bill, but it may also be taken to mean that if fishery matters were controlled by scientific men familiar with the natural history of the sea, and questions concerning fisheries were referred to marine biologists, recommendations would be made upon which reasonable regulations might be based. From *Nature* 3 May 1900.

50 YEARS AGO

The possibility of presenting television pictures of an adequate brightness on the large-size screen used in cinemas has been under development in Great Britain for some years. On April 29, the opportunity was taken by Messrs. Cinema-Television, Ltd., to demonstrate the state of this development by showing the B.B.C. Cup Final television programme to a selected audience of about a thousand persons, including television experts from some fifteen countries. The normal programme radiated from the Alexandra Palace transmitter was received in the neighbourhood of the Odeon Theatre, Penge, where the demonstration was given. The received signals were conveyed to special equipment placed in the auditorium of the theatre at a distance of some 12 metres from the screen. A very bright image of the television picture, about 16 cm. × 13 cm. in size, was formed on a special cathode-ray tube operating from a high-tension supply of 50 kilovolts. The optical projection system comprised a spherical mirror and plastic correcting plate by means of which the picture was thrown on to the theatre screen... In spite of the fact that, owing to weather conditions, the daylight at Wembley was on the dull side, the demonstration was very satisfactory. From *Nature* 6 May 1950.

elements of modern-human behaviour existed by then.

We can now add Walter and colleagues' discoveries⁴ to this picture. Middle Palaeolithic people might have spread from Africa along the shorelines of Arabia and into southern Asia during, or soon after, the last interglacial. Continuing along the narrow shorelines, to which they were already adapted, they could have progressed all the way to Indonesia at times of low sea level (Fig. 2). Movement along the coasts meant that they could have been spared the degree of habitat disruption faced by inland populations during the rapid climatic fluctuations of the Late Pleistocene. Coastal migration might also explain why they did not immediately replace archaic peoples living inland, such as those known from Ngandong in Indonesia¹. At what stage the coastal migrants first ventured out to sea is unknown, but from the Australian evidence it seems that it must have been before 60,000 years ago. Such behaviour may have developed through the need to ford rivers or extend coastal foraging areas.

Archaeologists might now concentrate profitably on exposed fossil beaches in regions such as Arabia and India. Such sites may well contain Middle Palaeolithic artefacts that further document the spread of modern humans and their adaptations to a coastal environment. Southern Asia must have formed an important secondary centre for dispersals of modern humans — there, too, it may have been the coasts that provided the first and fastest routes for migration, before movement inland up river valleys.

Chris Stringer is in the Human Origins Research Group, Department of Palaeontology, Natural History Museum, Cromwell Road, London SW7 5BD, UK.

1. Klein, R. *The Human Career* (Univ. Chicago Press, 1999).
2. Lahr, M. & Foley, R. *Yb. Phys. Anthropol.* **41**, 137–176 (1998).
3. Relethford, J. & Jorde, L. *Am. J. Phys. Anthropol.* **108**, 251–260 (1999).
4. Walter, R. C. *et al. Nature* **405**, 65–69 (2000).
5. Troeng, J. *Acta Archaeol. Lundensia* Ser. 8, No. 21 (1993).
6. Barton, R. N. *et al. Antiquity* **73**, 13–23 (1999).
7. Kingdon, J. *Self-Made Man and his Undoing* (Simon & Schuster, London, 1993).
8. Stringer, C. *Antiquity* **73**, 876–879 (1999).
9. Thorne, A. *et al. J. Hum. Evol.* **36**, 591–612 (1999).
10. *National Geographic* (map, Feb. 1997).

Fluid dynamics

Turbulence without inertia

Ronald G. Larson

In the 1880s Osborne Reynolds¹ established that fluid inertia (that is, momentum) drives the irregular patterns observed in water flowing rapidly from a pipe, plumes emerging from a smokestack, eddies in the wake of a bulky object, and many other everyday phenomena. Known as 'turbulence', these patterns occur at high values of the Reynolds number, the dimensionless ratio of inertial to viscous force. Over the years turbulence has become better characterized, and we now know it to be accompanied not only by an increase in drag, but also by certain characteristic spatial or temporal velocity fluctuations.

On page 53 of this issue, Groisman and Steinberg² show that both an increased resistance to flow and other features of turbulence can occur in fluids with hardly any inertial forces, if the role of inertia is instead played by elasticity, a force present in solutions of long-chain polymers. The flow studied by Groisman and Steinberg is simple: a polymer fluid confined between two parallel disks is sheared by rotation of one of the disks about their common axis. At increased flow rates (but rates still too low to generate much inertia) the flow acquires turbulent characteristics.

Irregular flow patterns have long been seen in polymeric and other elastic fluids, and have even occasionally been dubbed

'elastic turbulence'³. One such flow is that of a polymer melt emerging from the end of a capillary tube. Above a critical flow rate, the polymer jet becomes irregularly distorted (Fig. 1). Distortions in this and other polymer flows can mar products made from polymers, such as films and extruded parts. If the polymer is made more viscous by increasing the length of its polymer molecules, the minimum flow rate producing such irregular flow decreases. This is surprising because an increase in viscosity has the opposite effect on inertially driven turbulence. In fact, for fluids containing long polymer molecules, the Reynolds number at the onset of the instabilities can be minuscule, as low as 10^{-15} . Turbulent flow of water in a pipe, by contrast, has a much higher Reynolds number, around 10^5 .

It is clear, then, that inertia has nothing whatsoever to do with this kind of polymeric 'turbulence'. For viscous polymers, instabilities and irregular flows set in at a critical value of the so-called Weissenberg number, the ratio of elastic to viscous forces in the fluid, which for polymeric fluids plays the role of the Reynolds number in creating the nonlinearities that lead to unstable flow. Because elastic forces increase with polymer length more rapidly than do viscous forces, fluids containing long polymers are especially prone to such phenomena, despite having

viscosities that can be hundreds to millions of times higher than water. Until the work of Groisman and Steinberg², quantitative measurements of the length and time scales of elastic turbulence had been lacking.

Still lacking, even now, is a proper understanding of how the length and time scales of elastic turbulence are produced by the underlying elastic forces. In other words, what is missing is an elastic counterpart of the 'cascade' picture of inertial turbulence. In this, the largest eddies in the flow feed their energy into smaller eddies, which in turn drive even smaller eddies, and so on — until the energy stored in the smallest eddies is finally dissipated as heat. The resulting hierarchy of interacting eddies of various sizes is known as well-developed or hard turbulence. In Kolmogorov's 'bucket brigade' description of hard turbulence, a power-law distribution in length scales is produced by this handing off of energy from larger to smaller eddies⁴.

For 'elastic turbulence', the basic mechanisms of instability in the base flow have only recently been discovered. These mechanisms involve 'normal stresses'. Normal stresses are produced by the stretching of polymer molecules in a flow, leading to an elastic force like that in a stretched rubber band. If the streamlines are circular, the polymer 'rubber bands' press inward from all directions, generating a radial pressure gradient. This

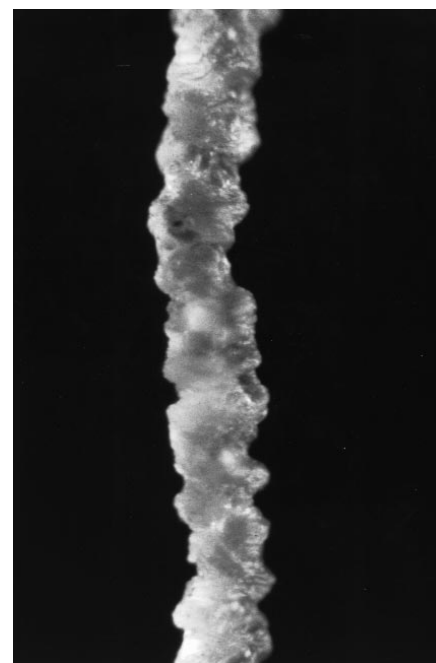


Figure 1 Is this jet turbulent? Turbulent flow is normally associated with a high Reynolds number. A polymer melt of high viscosity (5,000 pascal seconds) and low Reynolds number is greatly distorted after emerging from a capillary tube¹². Groisman and Steinberg² now confirm that the defining features of turbulence can appear in the flow of a polymer solution with high elasticity but low Reynolds number.

pressure, or 'hoop stress', leads to the famous 'rod climbing' phenomenon described by Weissenberg⁵ in 1947, and observed by every cook who has been annoyed by the climbing of cake batter (which contains natural polymers) up the shaft of an egg beater.

Despite being non-inertial, there is a strong analogy between such elastic forces and inertial instabilities involving curved streamlines, such as the flow in the gap between an inner rotating and a concentric outer stationary cylinder (circular Couette flow). In either the parallel-disk flow studied by Groisman and Steinberg, or the circular Couette flow, inertial forces try to fling fluid outwards, producing a radial pressure gradient, whereas elastic forces squeeze fluid inwards, tending to drive fluid up the rotating inner cylinder. This is what generates rod climbing. Whether the stresses are inertial or elastic (directed outward or inward), there is a radial pressure gradient that, if large enough, can drive an instability, leading to a more complex flow. For non-elastic fluids, the precise mechanism for the initial instability in inertial circular Couette flow was worked out by Taylor⁶ in 1923, whereas analogous work for purely elastic instabilities^{7–9} dates only from around 1985–95.

Beyond these initial instabilities, the first steps are being taken to work out the cascades of instabilities in elastic flows¹⁰. The work of Groisman and Steinberg leaps over these cascades, deliberately inciting highly unstable flow by having a large gap between the disks, relative to the disk diameter, so that the stabilizing influence of the viscous drag is minimized. The result is a flow with the power-law structures characteristic of hard, well-developed turbulence, but with negligible inertia.

Analogous 'hard elastic turbulence' should now be sought in other viscoelastic flows, such as Taylor–Couette, Taylor–Dean (curved pipe) flows, or the jet flow from a capillary. Or, it could be sought in structured fluids such as liquid crystals, where an ill-characterized irregular flow, sometimes called 'director turbulence', is observed¹¹. Equally important would be a detailed experimental characterization and theoretical explanation of the transition from the first instability in the base flow to the highly irregular patterns of elastic turbulence. In addition, it would be interesting to see how other well-known features of ordinary inertial turbulence, such as efficient fluid mixing, are reproduced in elastic turbulence, and whether these can be put to practical use. ■

Ronald G. Larson is in the Department of Chemical Engineering, University of Michigan, Ann Arbor, Michigan 48109-2136, USA.

e-mail: rlarson@engin.umich.edu

1. Reynolds, O. *Phil. Trans. R. Soc. Lond.* **174**, 935–982 (1883).
2. Groisman, A. & Steinberg, V. *Nature* **405**, 53–55 (2000).
3. Petrie, C. J. S. & Denn, M. M. *AIChE J.* **22**, 209–236 (1976).
4. Landau, L. D. & Lifshitz, E. M. *Fluid Mechanics* 2nd edn (Pergamon, Oxford, 1987).

5. Weissenberg, K. *Nature* **159**, 310–311 (1947).

6. Taylor, G. I. *Phil. Trans. R. Soc. Lond. A* **223**, 289–343 (1923).

7. Phan-Thien, N. J. *Non-Newton. Fluid Mech.* **13**, 325–340 (1983).

8. Larson, R. G., Shaqfeh, E. S. G. & Muller, S. J. *J. Fluid Mech.* **218**, 573–600 (1990).

9. Byars, J. A., Oztekin, A., Brown, R. A. & McKinley, G. H.

J. Fluid Mech. **271**, 173–218 (1994).

10. Khayat, R. E. *J. Fluid Mech.* **400**, 33–58 (1999).

11. Cladis, P. E. & van Saarloos, W. in *Solitons in Liquid Crystals* (ed. Lam, L.) 111–150 (Springer, New York, 1992).

12. Kalika, D. S. & Denn, M. M. *J. Rheol.* **31**, 815–834 (1987).

Apoptosis

Gone but not forgotten

Douglas R. Green and Helen M. Beere

Multicellular animals are confronted daily with death and its consequences, at least at the cellular level. Cells die from wear and tear, as a part of differentiation and selection, and through mechanisms that provide for normal cellular turnover. Cells usually die by apoptosis, a death process that is controlled by intrinsic cellular mechanisms. But, in cases of severe injury, cells may instead undergo necrosis (a 'passive' death resulting in cellular lysis). Under both circumstances, the dead cells are rapidly cleared from the body, but each leaves imprints of its passing that can have long-term consequences. Dead cells do tell tales, and one way in which they do so is described by Fadok and colleagues¹ on page 85 of this issue.

Apoptotic cell death was originally distinguished from necrosis on the basis of morphological differences and of the propensity for necrotic, but not apoptotic, cells to induce an inflammatory response. Although it is still unclear how necrotic cells trigger inflammation, this ability may be critical in linking tissue damage (and the consequent probable entry of harmful pathogens) to the generation of an immune response. In contrast, apoptosis is often described as a 'silent death'. The earthly remains of cells resulting from an apoptotic death are buried (or rather, 'eaten') by phagocytic cells, effectively eliminating all physical evidence of death. However, the physiological ghosts of this form of death persist in the form of altered behaviour of the phagocytes.

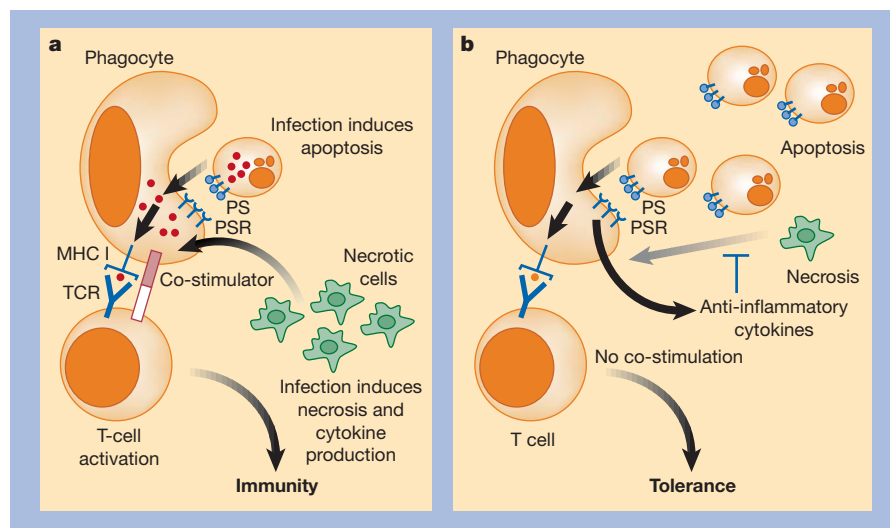


Figure 1 Phagocytes (including dendritic cells and macrophages) can engulf and degrade both apoptotic and necrotic cells. Apoptotic cells are taken up by means of their interaction with the phosphatidylserine (PS) receptor (PSR) on the phagocyte (as shown by Fadok *et al.*¹), as well as via other receptors (not shown). The balance between apoptosis and necrosis determines the biological response of the phagocyte. **a**, Under conditions of, for example, viral infection, necrosis and the release of pro-inflammatory cytokines are observed. Ingestion of the infected (apoptotic) cells by dendritic cells stimulates the presentation of viral antigenic peptides (red circles) by the class I major histocompatibility complex (MHC I) molecules. Necrotic cells are also ingested and, together with the pro-inflammatory cytokines, this process induces the expression of co-stimulatory molecules on the surface of the dendritic cell. Display of both of these markers stimulates T-cell activation and immunity. TCR, T-cell receptor. **b**, Under conditions of normal tissue turnover, apoptosis predominates. Ingestion of apoptotic cells by macrophages stimulates the release of anti-inflammatory cytokines and suppression of pro-inflammatory cytokines. Dendritic cells display peptides (small orange circle) from the engulfed apoptotic cell; however, expression of co-stimulatory molecules is not induced and T cells become tolerant to the displayed (self) peptide.

Recognition and removal of an apoptotic cell by macrophages — a type of phagocyte — is mediated by changes in the expression of membrane-associated markers on the dying cell². These markers selectively interact with specific receptors on the surface of the phagocyte. One common feature of late-stage apoptosis is exposure of the lipid phosphatidylserine, normally localized to the inner leaflet of the plasma membrane, on the outer-membrane leaflet of dying cells. This allows the uptake of dead cells by at least some types of phagocyte³.

It was assumed that a phosphatidylserine-specific receptor on the surface of the phagocyte must mediate these events, but candidate receptors failed to distinguish between phosphatidylserine and other lipids (something we know the phagocytes do). Now, Fadok and colleagues¹ have identified a phosphatidylserine receptor on the surface of activated human macrophages that selectively allows the phosphatidylserine-dependent uptake of apoptotic cells. Significantly, binding of phosphatidylserine to its receptor triggers the release of anti-inflammatory cytokines (such as transforming growth factor- β) and inhibits the production of pro-inflammatory cytokines (such as tumour necrosis factor- α), providing a link between the recognition of apoptotic cells and the physiological consequences of their uptake.

But the influence of dead cells on vertebrate physiology extends beyond the generation of pro- and anti-inflammatory cytokines⁴. The function of a phagocyte called a dendritic cell can be dramatically affected by both apoptotic and necrotic cells. However, the responses of the dendritic cell to these two stimuli are very different, and this may be one important basis for the ability of the immune system to respond to foreign, but not self, antigens^{5,6}.

Dendritic cells can engulf and degrade the proteins of target cells, which might be, for example, apoptotic cells resulting from normal tissue wear-and-tear, or necrotic cells following a viral infection (Fig. 1). The peptides generated from protein degradation are displayed on the surface of the dendritic cell by way of the major histocompatibility complex (MHC). These MHC-peptide complexes form the ligands for antigen receptors on T cells, which fight infection. In this way, MHC-peptide complexes regulate T-cell activity.

Uptake of necrotic cells by dendritic cells results not only in presentation of peptides on the cell surface, but also in activation of the dendritic cell to express co-stimulatory molecules⁷, which are necessary for T-cell activation and induction of an inflammatory response. In contrast, apoptotic cells do not trigger co-stimulator expression, even if they have undergone secondary necrosis following the induction of apoptosis⁷. Without the co-stimulatory molecules, the MHC-

X-ray astronomy

How far to the cosmic lighthouse?

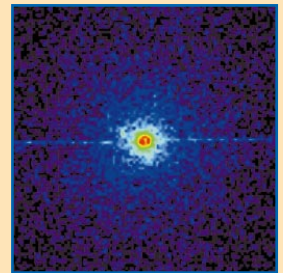
Determining distances to cosmic objects is a challenging business. Most techniques give only relative distances. X-rays detected by the Chandra space telescope from a star in our Galaxy provide a direct way to measure distances to galactic objects, and perhaps even other galaxies.

The star shown in red, Cygnus X-3, is a variable source of X-rays whose emission varies over a period of 4.8 hours. Interstellar dust scatters the X-rays by varying amounts, leading to a 'halo' of radiation. By measuring the difference in arrival times between radiation

from the inner halo (green ring) and outer halo (blue ring) a team of astronomers, led by Peter Predehl of the Max-Planck Institute in Garching, Germany, has worked out the distance to Cygnus X-3 to be 30,000 light years.

The uncertainty in the new measurement, to be published in the journal *Astronomy and Astrophysics*, is 20%. The accuracy was limited by the short observation time (3.5 hours) and the authors hope to improve this by observing the star over the whole of its 4.8-hour cycle.

The consequences of this



work may be far greater if the technique can be applied to variable X-ray sources in other galaxies. In particular, the distance to the Large Magellanic Cloud is crucial for estimating the age and expansion rate of the Universe, but there has to be enough dust around to produce a useful halo.

Sarah Tomlin

peptide complexes tend to selectively inactivate the T cells that recognize them⁶.

Dendritic cells that take up apoptotic cells process proteins present in the dead cells and shuttle peptide fragments into their class I MHC molecules, which are then displayed on the cell surface and can be recognized by T cells. This process of 'cross-priming' is remarkable in that, generally, the only peptides that are displayed on class I MHC molecules are those derived from the proteins present in that cell. Cross-priming defies that 'rule', because the dendritic cell presents peptides derived from the apoptotic cell that it has ingested⁸.

All of these results link together to offer an explanation for how the immune system discriminates between self and foreign proteins. During normal cell turnover, phagocytes (including dendritic cells) engulf the apoptotic cells and produce anti-inflammatory cytokines (for example, following engagement of the phosphatidylserine receptor, as shown by Fadok *et al.*¹). Any antigenic peptides presented, through cross-priming, on class I MHC molecules will fail to stimulate a T-cell response because of the absence of co-stimulation. Instead, the MHC complexes (which contain self peptides derived from the apoptotic cells of the organism itself) will selectively inactivate the appropriate T cells, ensuring that immune responses to these self proteins will be eliminated. Any potential low-level inflammation will be dampened by the induced release of anti-inflammatory cytokines¹, to make sure that no immune activation occurs.

But if, for example, viral infection occurs, the situation may be very different. An inflammatory response would be potentially induced as a consequence of extensive necrosis and by the cytokines (such as

interferons) produced by the infected cells. Under these circumstances, the antigens associated with class I MHC molecules and displayed on the surface of the dendritic cells would include peptides from both the necrotic cells (that is, self antigen) and viral proteins (that is, foreign antigen). Engulfment of necrotic cells and the pro-inflammatory cytokines act together to activate the dendritic cells, which, in turn, stimulate the production of specialized T cells (cytotoxic effector T cells). These cells eventually destroy the virally infected cells. Fortunately, responses to 'self' antigens do not occur, thanks to the inactivation of potentially responsive T cells during the previous 'quiet' periods of normal cellular turnover. So it appears that apoptosis ensures that inappropriate inflammatory responses to self-proteins do not occur.

The work of Fadok *et al.*¹ reinforces the idea that there are key connections linking the uptake of apoptotic and necrotic cells, the balance between immune homeostasis and inflammation, and activation versus tolerance induction in T cells. These mechanisms all function to ensure the efficient generation of specific immune responses when — and only when — they are needed. ■

Douglas R. Green and Helen M. Beere are at the La Jolla Institute for Allergy and Immunology, 10355 Science Center Drive, San Diego, California 92121, USA.

e-mail: dgreen5240@aol.com

1. Fadok, V. A. *et al.* *Nature* **405**, 85–90 (2000).
2. Ren, Y. & Savill, J. *Cell Death Differ.* **5**, 563–568 (1998).
3. Fadok, V. A. *et al.* *J. Immunol.* **149**, 4029–4035 (1992).
4. Fadok, V. A. *et al.* *J. Clin. Invest.* **101**, 890–898 (1998).
5. Gallucci, S., Lolkema, M. & Matzinger, P. *Nature Med.* **5**, 1249–1255 (1999).
6. Steinman, R. M., Turley, S., Mellman, I. & Inaba, K. *J. Exp. Med.* **191**, 411–416 (2000).
7. Sauter, B. *et al.* *J. Exp. Med.* **191**, 423–434 (2000).
8. Albert, M. L. *et al.* *J. Exp. Med.* **188**, 1359–1368 (1998).

Earth science

Strength of the San Andreas

Mark D. Zoback

Knowledge of the rates of motion and the long-term record of earthquakes along the San Andreas fault system in California has improved markedly over the past 25 years. Such knowledge has made possible long-term estimates of earthquake probability, typically for periods of 30 years¹. But our understanding of the physics of faulting remains woefully inadequate — questions regarding the feasibility of shorter-term predictions, and whether the occurrence of one earthquake increases the probability of an earthquake on another fault, remain unanswered.

Hence the interest in a provocative paper by Christopher Scholz, published in *Geology*². Scholz argues that the San Andreas fault (Fig. 1) is relatively 'strong'. Over the past three decades, hundreds of laboratory experiments have indicated remarkably similar friction coefficients of between 0.6 and 1.0 for a wide range of rocks³. Extrapolating these observations to natural faults implies that extremely high stresses should be required to cause fault motion at great depth, assuming normal (approximately hydrostatic) pore pressures. Pore pressure is the pressure of fluids, principally water, within the cracks and pores of a rock at depth. In the context of Scholz's argument, a strong San Andreas means that, to a depth of about 10–15 km (where large earthquakes nucleate), the frictional resistance to sliding along the fault is high. This is essentially consistent with the stresses predicted from laboratory friction experiments with hydrostatic pore pressures.

There seems little doubt that laboratory friction data are applicable to the faulted crust within Earth's relatively rigid lithospheric plates. Data from deep boreholes around the world show high levels of shear stress at depth, consistent with the laboratory friction experiments⁴. Scholz argues that the San Andreas (and by inference the other major faults which bound lithospheric plates) has essentially the same frictional strength as the faulted crust within the plates — in other words, that the hundreds of kilometres of offset that have occurred along the fault have been accomplished without significant weakening of the fault with respect to the surrounding brittle crust.

In the 1960s, stick-slip sliding was discovered in the laboratory as a possible analogue for crustal faulting⁵. Stick-slip is the process by which a fault in a rock being deformed at a steady rate in the laboratory would 'stick', allowing stress to build over time, and then suddenly 'slip' like an earthquake, releasing part of the accumulated

stress. Thus, the stick-slip phenomenon, along with the uniformity of laboratory friction coefficients, seemed to point directly to the applicability of laboratory friction experiments to earthquakes on major faults such as the San Andreas.

However, by the late 1960s it became clear from conductive heat-flow measurements in boreholes near the San Andreas that this rapid frictional shearing was not generating any significant heat^{6,7}. Plate tectonics had revealed that the San Andreas has slipped hundreds of kilometres at long-term average rates of several centimetres per year. If the San Andreas had the frictional strength predicted by the laboratory data (friction coefficients of 0.6–1.0), appreciable frictional heat would be generated by the long-term motion of the fault.

The lack of observable frictional heat implied that the average fault strength is a factor of approximately 5–10 lower than that implied by laboratory friction coefficients^{6,7}. Markedly lower fault strength could mean several things: that processes may be occurring during faulting that are not seen in laboratory experiments; that the rocks within the San Andreas are very different in composition to those typically found in exposed fault zones; or that fault slip is aided by fluids trapped within the fault zone at abnormally high pressure. One critical, but questionable, part of Scholz's argument is to discount the significance of the heat-flow measurements near the fault, even though these data show no evidence of frictionally generated heat or appreciable convective heat transport⁷.

A further factor to be taken into account is the discovery in the late 1980s that the direction of the maximum horizontal principal stress near the San Andreas is, in general, nearly perpendicular to the fault. This indicated that frictional sliding is occurring at levels of shear stress that are much lower than those existing in the immediately adjacent crust^{8,9}. Taken together, the heat flow and *in situ* stress orientation data indicated to many researchers that the San Andreas, and perhaps other plate-boundary faults, was a profoundly weak fault in an otherwise strong crust.

By assuming fault-normal compression at large distances from the San Andreas and stress-free slip on the fault at depths below the earthquake-generating zone, Scholz presents a model in which the direction of maximum stress in the upper crust rotates near the fault to a direction about 45° from the fault trend. Although one study suggests that such a crustal stress pattern may exist in the 'big bend' area of the San Andreas¹⁰ (Fig. 1), such stress rotations can be ruled out in the vicinity of the Loma Prieta¹¹ and Morgan Hill¹² earthquakes, as well as on the San Francisco peninsula¹³, where fault-normal compression is observed within 1–2 km of the fault trace. Thus, the stress rotation Scholz is modelling appears not to be a general characteristic of the San Andreas.

What makes the issue of fault strength even more controversial is that some researchers have argued, on the basis of geodynamical modelling, that Earth's brittle crust is generally weak, implying that both intraplate crust and plate-bounding faults such as the San Andreas have low strength¹⁴. So different groups of investigators now view plate-bounding faults such as the San Andreas as weak faults in a strong crust^{6–9}, weak faults in a weak crust¹⁴, or strong faults in a similarly strong crust².



Figure 1 The San Andreas fault. The fault marks the divide between the Pacific and North American tectonic plates, and accommodates about 35 mm of the annual 48 mm of relative motion between them. The northern and southern segments last broke in major earthquakes in 1906 (magnitude 7.8) and 1857 (magnitude 8.2), respectively; the central section creeps steadily and does not accumulate strain. The other places marked are cited in the text. The physics of faulting is poorly understood, and the question of whether the San Andreas is a weak fault in a strong crust, or a strong fault in a strong crust (as Scholz² proposes), remains controversial. (Courtesy of the US Geological Survey.)

So where does this leave us? The stress measurements from deep boreholes appear to rule out the second, weak-fault/weak-crust, hypothesis, at least for rigid plate interiors. Nonetheless, we still find ourselves faced with contradictory theories and data, and remain largely in the dark about the physics of faulting along the San Andreas and similar plate-bounding faults. The absolute strength of the San Andreas is likely to remain controversial until we obtain measurements, from within the fault itself, of the stress levels and physical conditions under which earthquakes occur. ■

Mark D. Zoback is in the Department of Geophysics, Stanford University, Stanford, California 94303, USA.

e-mail: zoback@pangea.stanford.edu

1. Working Group on California Earthquake Probabilities US Geol. Surv. Open File Rep. 99-517 (1999).
2. Scholz, C. *Geology* **28**, 163–166 (2000).
3. Byerlee, J. D. *Pure Appl. Geophys.* **116**, 615–626 (1978).
4. Townend, J. & Zoback, M. D. *Geology* **28**, 399–403 (2000).
5. Brace, W. F. & Byerlee, J. D. *Science* **153**, 990–992 (1966).
6. Brune, J. N., Henyey, T. L. & Roy, R. F. *J. Geophys. Res.* **74**, 3821–3827 (1969).
7. Lachenbruch, A. H. & Sass, J. H. *J. Geophys. Res.* **97**, 4995–5015 (1992).
8. Zoback, M. D. *et al. Science* **238**, 1105–1111 (1987).
9. Mount, V. S. & Suppe, J. *Geology* **15**, 1143–1146 (1987).
10. Hardebeck, J. & Hauksson, E. *Science* **278**, 650–653 (1999).
11. Zoback, M. D. & Beroza, G. *Geology* **21**, 181–185 (1993).
12. Oppenheimer, D. H., Reasenberg, P. A. & Simpson, R. W. *J. Geophys. Res.* **93**, 9007–9026 (1988).
13. Zoback, M. L., Jachens, R. C. & Olsen, J. A. *J. Geophys. Res.* **104**, 10719–10742 (1999).
14. Molnar, P. in *Fault Mechanics and Transport Properties of Rocks* (eds Evans, B. & Wong, T.-F.) 435–459 (Academic, London, 1992).

Molecular evolution

A single birth of all plastids?

Jeffrey D. Palmer

Amidst the present swirl of uncertainty about the correct branching order of the tree of life — or whether gene transfer between species on different branches is so pervasive as to make a coherent tree unattainable — a few conclusions about early evolution stand out as unassailable. Foremost among them is that plastids (the organelles in which photosynthesis occurs) and mitochondria (the cellular power plant, where respiration takes place) arose aeons ago from bacteria that set up home within other unicellular organisms, to mutual benefit. Only a die-hard creationist, by denying evolution itself, would deny that plastids and mitochondria are highly derived bacteria, with most of their genes having been either lost or transferred to the nucleus of the host cell. Although it is firmly established that plastids originated from cyanobacteria (Fig. 1), the number of such origination events is in dispute.

Primary plastids are the double-membrane-bound products of a symbiotic (mutually beneficial) relationship between eukaryotic cells (those with a nucleus) and cyanobacteria. Such plastids are found in three distinctive groups of eukaryotes — the green algae (including land plants), the red algae, and the glaucophytes (Fig. 1). Accumulating evidence from plastid^{1–3} and mitochondrial⁴ genomes favours a single (monophyletic) origin of the primary plastids, but the information from nuclear genes is at best equivocal^{1,5}. On page 69 of this issue, however, Moreira and colleagues⁶ report the first strong evidence from nuclear genes for a common origin of green- and red-algal plastids, and weaker evidence that glaucophyte plastids arose by the same event. These results are an important step towards universal acceptance of a monophyletic origin of plastids.

The study by Moreira *et al.*⁶ provides three mutually reinforcing lines of evidence for an affiliation of green and red algae. First, they analysed the evolutionary relationships (phylogeny) of nuclear genes encoding elongation factor-2, a protein involved in the

translation of RNA into protein. The nuclear sequences were from a variety of species, and included newly determined sequences from red algae and protists (single-celled eukaryotes). The results strongly support the idea that green and red algae are ‘sister groups’, or closest relatives. Second, Moreira *et al.* analysed the gene encoding subunit RBP1 of RNA polymerase. Previous studies of this molecule resulted in a rejection of the possibility of a green algal/red algal clade⁵ (a clade is a group of species that share a common ancestor). But Moreira *et al.*’s analysis included more taxa and shows that this clade can no longer be rejected on such grounds. Third, analysis of a combined data set of 13 nuclear proteins provides strong support for the green algal/red algal clade. In addition, a six-gene analysis identified glaucophytes as the sister group of the green/red clade, but with only weak support.

These results are consistent with a single origin of all primary plastids; with Cavalier-Smith’s proposal⁷ of a ‘Kingdom Plantae’ composed of green algae, red algae and glaucophytes; and with evidence from the phylogeny of plastid genes⁸, and from the retention of a bacterial cell wall by glaucophyte plastids, that glaucophytes represent the first branch of plastid evolution.

Two other studies of nuclear genes support a monophyletic origin of plastids. Van

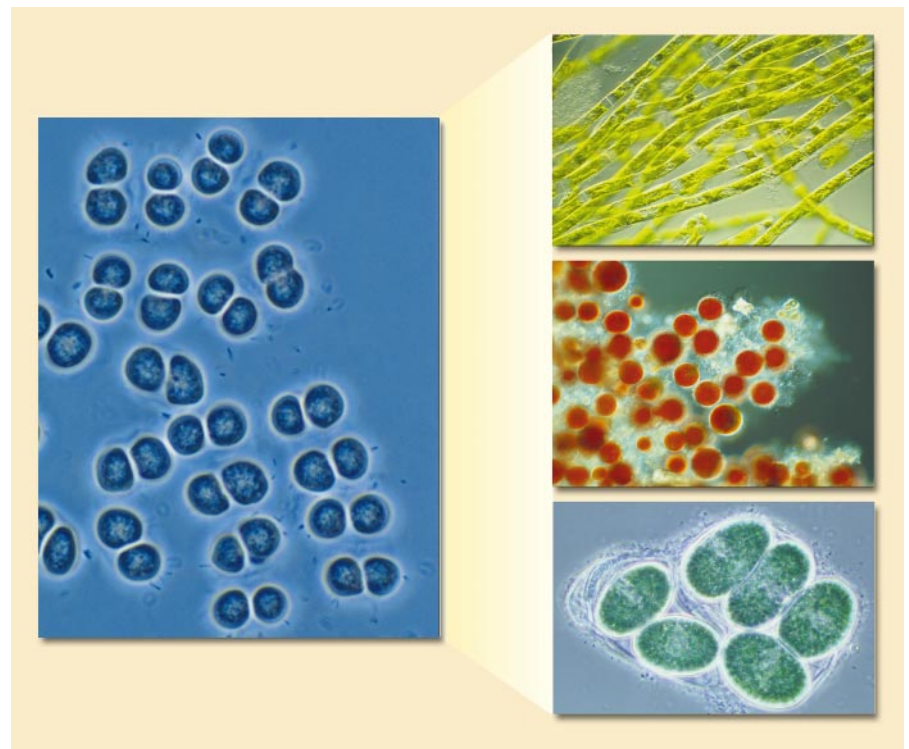


Figure 1 A single origin of plastids. Cyanobacteria (left, represented here by *Eucaupsis* sp.; magnification $\times 745$) probably gave rise to plastids of green algae (filaments of *Ulothrix fimbriata*; magnification $\times 120$), red algae (*Porphyridium* sp.; magnification $\times 77$), and glaucophytes (*Glaucocystis* sp.; magnification $\times 240$) through a single symbiotic event that took place over 1.2 billion years ago. This theory is supported by the work of Moreira *et al.*⁶. The photograph of *P. elegans* was provided by Astrid and Hans-Frieder Michler/Science Photo Library. The other photographs were provided by Michael Abbey/Science Photo Library.

der Auwera *et al.*⁹ have done a combined analysis of small- and large-subunit ribosomal RNAs. Their results are important because — unlike the poorly resolved trees based on small-subunit rRNAs that are sometimes cited as evidence against plastid monophyly — their 'rate-calibrated' tree finds strong support for the green/red clade, albeit with limited sampling of the major groups of eukaryotes (and no glaucophytes). A four-gene study by S. L. Baldauf *et al.* (personal communication) features more impressive sampling of eukaryotes and reveals, although with low support, a monophyletic assemblage of all three primary-plastid-containing groups.

How far are we, then, from being able to write with textbook-like certainty of a single cyanobacterial origin of all plastids? We need more information from both mitochondrial and nuclear genes of glaucophytes. Fortunately, sequencing of the mitochondrial genome of the glaucophyte *Cyanophora paradoxa* is nearing completion. Preliminary analyses (F. Lang, personal communication) place it, with modest support, as the sister group to the strongly supported green algal/red algal clade defined by previous analysis of mitochondrial genomes⁴.

For most relevant sets of data from nuclear and mitochondrial genes, sampling within eukaryotes is rather poor, and in most plastid data sets the cyanobacteria are poorly represented. This increases the chances that phylogenetic artefacts may have led to spurious placements of taxa in these poorly sampled trees. As Moreira *et al.*⁶ point out, sampling is a particular problem for their multigene analyses, which include just one lineage (animals plus fungi) from which many eukaryotes were studied, and only six or seven groups from which just one eukaryote was sampled. A particular concern is whether the strong support for the green algal/red algal clade obtained in the 13-protein analysis arises largely from inclusion of elongation factor-2 — the only molecule that on its own provided strong support for this group. One wonders what a 12-protein analysis (that is, excluding elongation factor-2) would show.

These caveats aside, it is nonetheless impressive that so many phylogenies constructed from analysis of all three genomes — as well as derived commonalities of plastid targeting, gene content and arrangement, and the composition of light-harvesting pigments^{1–3,10} — are painting the same picture. That is, the primary plastids in red and green algae and glaucophytes all arose from a single, ancient symbiosis between a cyanobacterium and a eukaryote. We are rapidly approaching the almost untestable proposition that two or more symbiotic events could have taken place only if they occurred in a relatively short period and only if they involved closely related groups of eukaryotic hosts

and cyanobacterial guests, both presumably well adapted for their respective roles.

As Cavalier-Smith has argued⁷, among the implications of a single origin of primary plastids is a single origin of the complex machinery by which plastids import proteins made in the main body of the cell. In support of this idea, transit peptides from red algae and glaucophytes direct efficient import of proteins into plastids of land plants, and vice versa (for glaucophytes at least)¹. A second implication is that the variety of light-responsive pigments of primary (and secondary) plastids probably reflects differential loss of one or more such pigments and associated proteins from a cyanobacterial ancestor that was fully equipped with all of these compounds¹¹.

When did the breakthrough symbiosis occur? Cavalier-Smith⁷ speculates that it was as recent as 600 million years ago, but this seems increasingly unlikely given the evidence for red algal fossils of 1.2 billion years in age¹². Fossils of controversially algal origin date to 2.1 billion years¹³, and the first cyanobacteria date back at least 2.7 billion years¹⁴. What did this cyanobacterial ancestor look like, and did it possess any of the gene arrangements and types of light-harvesting complex that are thought to be defining features of (monophyletic) plastid evolution^{1–3,10}? Sadly, phylogenetic analyses have been unable to identify specific relatives of plastids among living cyanobacteria, making this problem hard to tackle.

A single origin of primary plastids now seems certain. But the number of secondary symbioses (that is, ingestion by a host cell of a eukaryotic cell that already contains plastids) is controversial. Estimates range from as few as two⁷ to as many as seven¹. Answering this question will require not only — as with primary symbiosis — an improved understanding of phylogeny, but also a search for symbiotically derived genes in eukaryotes postulated⁷ to have lost plastids of secondary origin. ■

Jeffrey D. Palmer is in the Department of Biology, Indiana University, Bloomington, Indiana 47405, USA.

e-mail: jpalmer@bio.indiana.edu

1. Palmer, J. D. & Delwiche, C. F. *Plant Syst. Evol.* **11** (suppl.), 53–86 (1997).
2. Stoebe, B. & Kowallik, K. V. *Trends Genet.* **15**, 344–347 (1999).
3. Turmel, M., Otis, C. & Lemieux, C. *Proc. Natl Acad. Sci. USA* **96**, 10248–10253 (1999).
4. Burger, G., Saint-Louis, D., Gray, M. W. & Lang, B. F. *Plant Cell* **11**, 1675–1694 (1999).
5. Stiller, J. W. & Hall, B. D. *Proc. Natl Acad. Sci. USA* **94**, 4520–4525 (1997).
6. Moreira, D., Le Guyader, H. & Philippe, H. *Nature* **405**, 69–72 (2000).
7. Cavalier-Smith, T. *Trends Plant Sci.* **5**, 174–182 (2000).
8. Martin, W. *et al. Nature* **393**, 162–165 (1998).
9. Van der Auwera, G., Hofmann, C. J. B., De Rijk, P. & De Wachter, R. *Mol. Phylogenet. Evol.* **19**, 333–342 (1998).
10. Durnford, D. J. *J. Mol. Evol.* **48**, 59–68 (1999).
11. Tomitani, A. *et al. Nature* **400**, 159–162 (1999).
12. Butterfield, N. J. *Palaeobiologica* (in the press).
13. Han, T. M. & Runnegar, B. *Science* **257**, 232–235 (1992).
14. Brock, J. J., Logan, G. A., Buick, R. & Summons, R. E. *Science* **285**, 1033–1036 (1999).

Daedalus

A flicker of interest

The human eye is sensitive to change above all else. Hence the flashing indicator for cars, the blink comparator and the flicker photometer, and the blinking cursor on the computer screen. Indeed, says Daedalus, hence the strange and terrible authority of the computer screen itself, and of the TV and cinema screen as well. The subliminal flicker of these hypnotic displays rivets the attention and subverts the mind.

So Daedalus wants to give printed text and pictures, and indeed objects in the real world, the eye-appeal of constant flicker. What is needed is a dye whose reflectivity is not steady, but oscillates about a mean value. Ordinary phosphors absorb light and emit it later. The photon energy is stored in the crystal structure as excited unpaired electrons. So Daedalus is inventing an oscillating phosphor.

He recalls that unpaired electrons are paramagnetic. An excited state should therefore induce a local magnetic field; and the rate at which it decays should depend on the local field. DREADCO solid-state physicists are therefore growing tiny phosphor crystals each containing a two-state magnetic inclusion. The idea is that the phosphor will absorb light until its growing field 'flips' the magnetic element. This will trigger the rapid re-emission of light; the field will decay and the magnetic inclusion will flip back. The result will be an unsteady, oscillating phosphor, driven by the energy of the light falling on it. It will brighten and darken endlessly about its mean reflectivity.

Artists and ad-men will rush to exploit the new products. DREADCO's 'Wink' inks and paints will flicker beckoningly from posters, magazines and the more garish of fashion clothes and accessories. Warning signs, safety regulations, identifying marks and disclaimers in contracts will signal their importance in unmissable Wink. High-frequency Winks, compelling subliminal attention without it being obvious why, will revolutionize the cosmetics industry. Its customers, without seeming garish or blatant, will now be able to call subtle attention to their most charming features. Long-period Winks will alleviate the visual boredom of institutions and offices; they will even vary in frequency with the ambient illumination. In theatres and clubs, strobe lighting on Winked decor will induce amazing interference effects. Sadly, patrons may be at risk of epileptic fits. **David Jones**

The Further Inventions of Daedalus is published by Oxford University Press.

Obituary

Jane Gray (1929–2000)

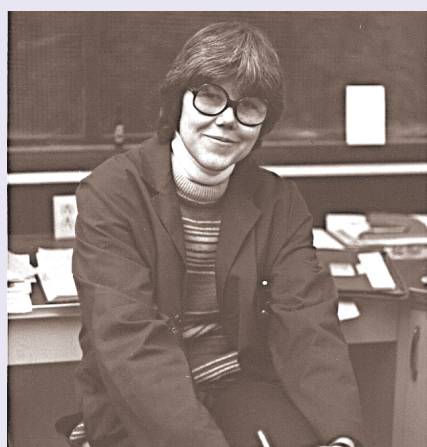
Jane Gray, a controversial figure in palaeoecology, died of acute liver failure caused by cancer on 9 January 2000. She was especially known for proposing a startling new interpretation of the invasion of land by life, a view that was at first almost universally resisted but which has now become a part of the standard view. Gray was a tenacious fighter for her ideas and so was regarded as 'difficult' by many people. But those who knew her best found her to be a generous, original and acutely intelligent colleague.

I first met Jane Gray in 1983, at a meeting of the American Association for the Advancement of Science, where she had helped organize a symposium entitled "The Greening of the Continents". I was in attendance as a novice palaeontologist (having spent a few decades as a systematic biologist before being attracted to the study of fossils). Gray and Arthur Boucot presented their evidence, based on microfossils including spores, for the presence of plants on land in the Caradocian epoch (Ordovician period), now dated at 458–449 million years ago.

Her presentation met with strong opposition from the formidable Harlan Banks, William Chaloner and others, and the discussion became somewhat heated. I remember whispering to Ian Rolfe (in jest), "These people are ready to kill over spores!". But I came away with the impression of Gray as fully capable of responding, with a withering command of the evidence, to any challenge.

Gray's personal code, intolerant of misbehaviour professional or otherwise, was undoubtedly formed by her family background, with its roots in the military establishment. Both her father (Colonel E. B. Gray) and uncle were West Pointers. She herself was educated in geology at Radcliffe College, Massachusetts (now part of Harvard University), and was introduced to palaeobotany by Elso Barghoorn. Most of her future work would be in palynology, the study of pollen systematics, especially from geological deposits. But Gray always considered herself a palaeoecologist, although she had an acute eye for systematic detail and a full command of spore taxonomy.

In 1952, Gray became one of the first recipients of a National Science Foundation Fellowship for study abroad, and sharpened her skills in pollen analysis with Johannes Iversen in Denmark. Her PhD came in 1958 from the University



Palaeoecologist who specialized in study of the colonization of land by plants

of California at Berkeley, under palaeobotanist Ralph Chaney, and she retained a life-long loyalty to that institution.

The 1950s and 1960s were not particularly female-friendly times in science, and Gray's uncompromising personality and readiness to defend her work made her rise through the academic ranks difficult. 'Anti-nepotism' regulations made it impossible for Jane and her husband, Antone Jacobson, to work at the same institution. Long periods of separation eventually led to a divorce.

After her move to the University of Oregon in 1963, Gray had worked almost entirely on problems in the palaeoecology of the Tertiary period (roughly 65–1.6 million years ago). But on his arrival in 1970, Boucot urged her to begin investigating the early invasion of the continents by plants. Her first evidence pushed the date back to the early Silurian period, about 430 million years ago, and later, collaborating with Boucot, to the Caradocian. Her most recent work, on spores from Arabia, pushed the date back a further 10–12 million years. The opposition of the palaeobotanical establishment made funding difficult to obtain. But Gray was as astute at playing the stock market as she was at interpreting fossil spores, and used her independent wealth to fund her own research.

In Boucot, a palaeozoologist and specialist in marine invertebrates known as brachiopods, Gray found an ideal collaborator, and together they began to

explore early palaeoecology from an eclectic viewpoint. Gray and Martha Sherwood-Pike discovered the earliest terrestrial fungi, and Gray and Boucot possibly the earliest terrestrial animal remains in the early Silurian Tuscarora Formation in Pennsylvania. Her interests continued to be diverse, and she wrote on topics ranging from redwood trees to the unicellular euglenoids. A signal accomplishment in 1988 was an exhaustive book, *Paleolimnology: Aspects of Freshwater Paleoeecology and Biogeography*, published by Elsevier. Seminal work on early animal remains appeared in 1993, in an article that took off from the Tuscarora fossils to outline an evolutionary framework for the development of the animal component of the first terrestrial ecosystems.

One of Gray's most lasting contributions was the creation of a spore-zonation framework for land plants of the Ordovician–Silurian, based on size and ornamentation of spores, and their occurrence in tetrads, dyads and as dispersed single spores. In her interpretation, this work documented the successive appearance of liverworts, mosses and vascular plants (or their ancestors) on land.

My own brief collaboration with Gray was rewarding. Her way of looking at old evidence anew, and her editorial acuity, helped us produce a widely reprinted article on the evolution of terrestrial systems. My last encounter with Gray, at a meeting of the Geological Society of America, was delightful as always, but raised concern for her health — she had suffered for many years from diabetes, and her liver problems had begun to emerge.

Obviously weaker physically, she remained mentally sharp. Indeed, the many projects in train at the time of her death showed her interests to be even broader than before. Still to appear are an article on controls on Earth's atmosphere before about 400 million years ago, a monograph that will question current models of the carbon dioxide and oxygen content of that early atmosphere. A huge manuscript, *The Evolution of the Nonmarine Ecosystem*, is only partially completed. Gray's massive collection will be moved to the University of Sheffield in Britain. Her works, published and unpublished, constitute a rich palaeoecological legacy of fact and theory.

William A. Shear

William A. Shear is in the Department of Biology, Hampden-Sydney College, Hampden-Sydney, Virginia 23948, USA.
e-mail: bills@hsc.edu

Beetle larvae cooperate to mimic bees

These parasites get into bees' nests by fooling males into trying to mate with them.

The life cycles of parasites often involve complex behavioural and morphological adaptations in order to find a host. Here we report a remarkable mode of host-finding by the blister beetle *Meloe franciscanus*, in which young larvae aggregate together on vegetation to mimic the appearance of a female bee, luring male bees to land on them and collect the aggregation as a unit for transmitting to females during real matings. Although cooperative behaviour is common among highly social insects, particularly bees^{1–3}, to our knowledge it has not been reported before in blister beetles, nor has it been associated with mimicry.

Although some mites⁴ also parasitize their hosts by venereal transmission, this behaviour is not seen in other *Meloe* species, whose larvae disperse to flowers and attach individually to passing bees whose nests they parasitize. Young *Meloe* larvae (triungulins) are highly adapted for grasping bees, structurally resembling lice rather than beetle larvae (so much so, that Linnaeus described a triungulin as a louse rather than a beetle⁵). *Meloe* triungulins drop off when they arrive at the bee's nest, however, and develop primarily on pollen provided by the bee^{6–8}.

We investigated the interaction of bees and triungulin aggregations at the Kelso Dunes in California's Mojave Desert during April 1992 and April and May 1999. Triungulins cooperated in forming aggregations and in holding onto vegetation (Fig. 1a). They collectively responded to outside stimuli, such as nearby movements, by waving their front legs or by contracting as a unit. They also moved as a unit, sometimes forming living bridges between adjacent grass blades. Long-lasting aggregations (mean, 5.4 days; $n=36$; range, 1–15 days) were found on grass and twigs, but there were no aggregations or individual triungulins on or near 1,256 flowers inspected in 1999.

Activities of the anthophorid bee *Habropoda pallida* were positively associated with the aggregations. We observed 98 instances in which bees hovered within a few centimetres of an aggregation and nine instances in which bees landed on an aggregation. These bees often flew away with all or most of an aggregation attached to their ventral surface. All of the bees that were orientated towards masses and could be sexed ($n=42$) were male. We determined the fate of 22 of the triungulin aggregations: of these, five were removed by a bee, seven were killed by adverse weather conditions, nine died of old age

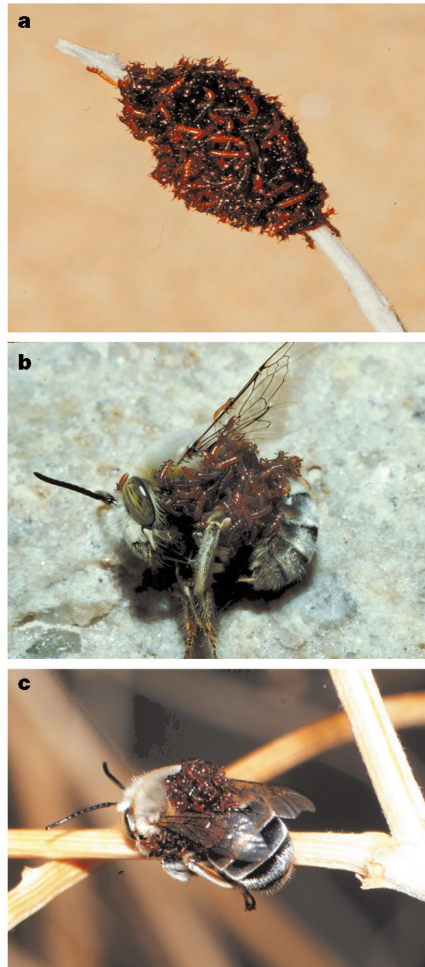


Figure 1 Interaction of *Meloe franciscanus* triungulin aggregations and the bee *Habropoda pallida*. **a**, Triungulins aggregated on a dried twig. Aggregations sampled in 1992 included 120–2,359 larvae (mean, 549; $n=19$; s.d. = 460) and crudely resembled perching bees in shape, as well as their abdomen colour. From 17 April to 18 May 1999, we monitored 36 triungulin aggregations ranging in size from 1.3 to 0.3 cm from 08:00 to 17:00 each day. We concentrated observations on a 10 m × 6 m portion of an area that contained eight aggregations. In addition, one of us periodically checked the status of aggregations in other areas and monitored the activity associated with them. Although aggregations appear reddish to humans, bees are red-blind and confuse red with black¹⁰. Triungulin aggregations and models, as well as bees and *Astragalus lentiginos* var. *borreganus* flowers, were photographed with a Kodak 18 A filter to detect ultraviolet reflectance patterns to which bees might respond. Only flowers showed an ultraviolet pattern different from that at visible wavelengths. **b**, Male bee with triungulins on its underside. In 1992, we collected male bees from sleeping aggregations and as bees visited *Astragalus* flowers. We tested the null hypothesis that triungulins had attached to bees at random by computing the index of dispersion, I_b , of triungulins on bees: $I_b = s^2(n-1)/\bar{X}$, which is approximately distributed as χ^2 with $n-1$ degrees of freedom¹¹. Triungulins were highly aggregated on male bees (mean, 53 triungulins per bee; $n=26$, s.d. = 80; range, 4–369; $I_b = 6,100$, 23 d.f., $P < 0.0001$). **c**, Female bee with triungulins on its dorsal surface. The abdomen of females and the aggregations are similar in size (mean, 0.69 cm; $n=6$ for abdomens; mean, 0.63 cm; $n=29$ for aggregations).

and one was eaten by a spider. In addition, ten masses disappeared in a way consistent with removal by bees. We estimate that bees removed 42% of the masses.

All male bees sampled in 1992 carried triungulins (Fig. 1b). Most (26 of 27) carried them on their underside, where they would be transferred to the dorsal surface of a female during attempted matings in which males typically mount females. All females that we observed with triungulins ($n=7$) carried them on their dorsal surface (Fig. 1c), and we saw a female that apparently acquired newly deposited triungulins immediately after copulating.

We believe that male bees confuse triungulin masses with female bees, which often perch on vegetation in positions similar to those presented by the masses, because males approach and land on masses and females in the same way. This phenomenon could be mediated in a way similar to pseudocopulation in orchids⁹ and probably involves both visual and olfactory mimicry. The importance of olfactory cues is sup-

ported by our observation of four males landing on or hovering next to groups of triungulins before their formation into discrete aggregations, and by the fact that bees ignore painted models of aggregations placed nearby.

Females of *Meloe* species produce up to 3,000 larvae per clutch⁶. Such a high reproductive rate is usually associated with high larval mortality and low individual success in host-finding. We believe that the mimicry practised by *M. franciscanus* larvae, together with their subsequent venereal transmission, enhances their chance of finding a bee's nest. This may apply particularly in highly variable environments such as our desert study site, where bee numbers and the location and frequency of flowering plants vary widely from year to year.

John Hafernik, Leslie Saul-Gershenz

Department of Biology,
San Francisco State University,
1600 Holloway Avenue, San Francisco,
California 94132, USA
e-mail: hafernik@sfsu.edu

1. Wilson, E. O. *The Insect Societies* (Harvard Univ. Press, Cambridge, MA, 1971).
2. Visscher, P. K. & Camazine, S. *Nature* **397**, 400 (1999).
3. Crespi, B. J. & Choe, J. C. in *The Evolution of Social Behavior in Insects and Arachnids* (eds Choe, J. C. & Crespi, B. J.) 499–524 (Cambridge Univ. Press, 1997).
4. Cooper, K. W. *Trans. Am. Entomol. Soc.* **80**, 119–174 (1955).
5. Linnaeus, C. *Systema Naturae. Regnum Animale* 12th edn (Salvii, Stockholm, 1767).

Ecosystems

Coral bleach-out in Belize

The highest sea surface temperatures ever recorded, related both to the 1997–98 El Niño/Southern Oscillation and to global warming¹, caused severe bleaching of corals worldwide in 1998 (ref. 2). This thermal anomaly induced mass mortality of scleractinian corals on lagoonal reefs in Belize, the first time that a coral population in the Caribbean has collapsed completely from bleaching. Cores extracted from the Belizean reefs showed that these events were unprecedented over at least the past 3,000 years.

Coral bleaching occurs when coral colonies under physiological stress expel their symbiotic algae (zooxanthellae). Usually a response to high temperature and/or high incident solar radiation, bleaching has become more frequent on coral reefs worldwide over the past two decades^{3,4}. Bleaching-related mass coral mortalities, involving population collapses that can lead to local extirpation, have occurred several times recently in the Indo-Pacific^{5,6}. In contrast, bleaching in the Caribbean has until now been followed by substantial recovery of the affected coral populations⁷.

There are no records of mass bleaching along the Belizean barrier reef before an episode in 1995, from which most coral colonies recovered⁸. In 1998, sea temperatures in the central sector of the barrier reef, which rarely exceed 29 °C, were greater than 30 °C (maximum, 31.5 °C) at depths of 2 and 10 m from 10 August to 14 October (K. H. Koltes, personal communication). Storm waves and rain associated with hurricane Mitch (27–31 October) reduced temperatures by 1 °C or less, but temperatures did not drop below 29 °C until 6 November. Mass bleaching was evident in both fore-reef and lagoonal environments during the summer and autumn (T. Bright, personal communication). Some coral colonies on the fore reef experienced partial mortality, but most colonies recovered their coloration in the months following hurricane Mitch⁹. In contrast, levels of coral mortality were catastrophic in the lagoon.

The lettuce coral *Agaricia tenuifolia* was the most abundant coral on reefs of the central Belizean lagoon in 1998 (Fig. 1).

6. Pinto, J. D. & Selander, R. B. *Biol. Monogr.* **42**, 1–222 (1970).
7. Erickson, E. H. & Werner, F. G. *Ann. Entomol. Soc. Am.* **67**, 903–908 (1974).
8. Clausen, C. P. *Annu. Rev. Entomol.* **21**, 343–368 (1976).
9. Borg-Karlson, A. K. *Phytochemistry* **29**, 1359–1387 (1990).
10. von Frisch, K. *Bees, their Vision, Chemical Senses, and Language* (Cornell Univ. Press, Ithaca, NY, 1950).
11. Southwood, T. R. E. *Ecological Methods* (Chapman & Hall, London, 1978).

Surveys at two sites on 22–23 October showed that virtually all living coral colonies were bleached white. Some of the bleached *Ag. tenuifolia* had already died by late October (Fig. 1b); the dead skeletons were standing in their growth position, indicative of recent mortality. Complete bleaching was also evident in almost all large, plate- and head-forming species down to the lagoon floor at 21 m. Surveys in 1999 and 2000 revealed that *Ag. tenuifolia* had undergone almost total mortality at all

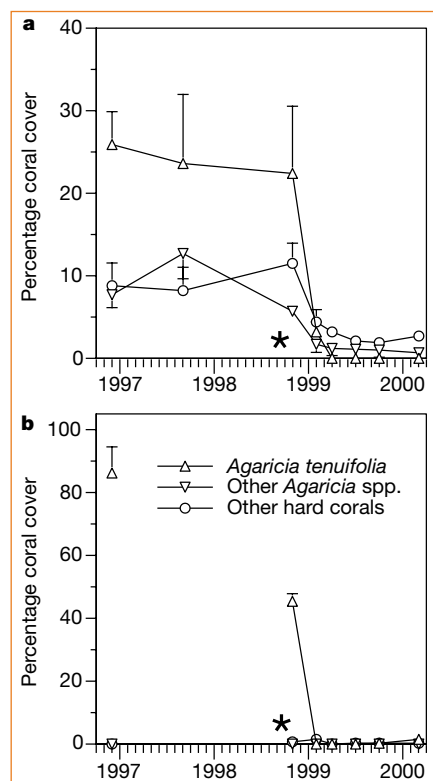


Figure 1 Changes in coral cover from December 1998 to February 2000 on two lagoonal reefs. **a**, Changes at Channel Cay, and **b**, at Cat Cay, which is 4 km away (~16° 40' N, 88° 10' W). Error bars represent standard errors. Asterisks indicate the onset of the high-temperature anomaly in August 1998. Percentage cover was estimated by laying a measuring tape down the reef slope along permanent transects and recording the sessile organisms every 10 cm. Hard corals include Scleractinia (true corals) and Milleporina (fire corals); the latter always constituted much less than 1% cover. Transects were about 20 m long and spanned 3–15 m depth at Channel Cay ($n=3$); transects were about 8 m long and spanned 3–8 m depth at Cat Cay ($n=5$). Surveys were made in December 1996, August 1997, October 1998, in January, March, June and October 1999, and in February 2000. Virtually all living colonies in October 1998 were completely bleached. The subsequent decline in coral cover was significant at $P=0.004$ (sign test).

depths. Heavy mortality occurred in the other species as well (Fig. 1a). Similar patterns were seen throughout a 375-km² area of the central lagoon. High water turbidity and the broad depth range of these effects suggest that temperature, rather than solar radiation, was primarily responsible for the bleaching and subsequent coral mortality.

Ag. tenuifolia had been the dominant space-occupier on lagoonal reefs since the late 1980s. At that time, the previous dominant, *Acropora cervicornis* (staghorn coral), was almost eliminated by white-band disease (a presumed bacterial infection unrelated to bleaching), permitting *Ag. tenuifolia* to increase opportunistically¹⁰. To determine whether similar events occurred in the past, we extracted 12 cores by hand from nine reefs in the central lagoon.

These push-cores penetrated several metres into the uncemented reef frameworks. Radiocarbon analysis yielded a maximum uncorrected ¹⁴C date of more than 3,000 yr. The cores showed continuous domination of lagoonal assemblages by *Ac. cervicornis*, with no widespread *Acropora* mortality, and no replacement by and subsequent disappearance of *Ag. tenuifolia*, in the three millennia before the late 1980s. The loss of *Ac. cervicornis* to disease, its replacement by *Ag. tenuifolia*, and the loss of *Agaricia* to bleaching a decade later were novel events on a timescale of millennia, and they occurred following an extended period of stable sea level (a sea-level rise of about 1 m in the past 3,000 yr)^{10,11}.

There is growing concern that global climate change is degrading coral reef ecosystems^{12,13}, with coral mortality increasing as a result of bleaching and emergent diseases^{4,12,14}: our results from Belize appear to justify this concern.

Richard B. Aronson*, **William F. Precht†**, **Ian G. Macintyre‡**, **Thaddeus J. T. Murdoch***

*Dauphin Island Sea Lab, 101 Bienville Boulevard, Dauphin Island, Alabama 36528, USA

e-mail: raronson@jaguar1.usouthal.edu

†PBS&J, 2001 Northwest 107th Avenue, Miami, Florida 33172, USA

‡Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, Washington DC 20560, USA

1. Hansen, J. et al. *J. Geophys. Res.* **104**, 30997–31022 (1999).
2. International Society for Reef Studies *Reef Encounter* **24**, 19–20 (1998).
3. Glynn, P. W. *Coral Reefs* **12**, 1–17 (1993).
4. Hoegh-Guldberg, O. *Mar. Freshwat. Res.* **50**, 839–866 (1999).
5. Glynn, P. W. in *Global Ecological Consequences of the 1982–83 El Niño–Southern Oscillation* (ed. Glynn, P. W.) 55–126 (Elsevier, Amsterdam, 1990).
6. Oliver, J. *Proc. 5th Int. Coral Reef Congr.* **4**, 201–206 (1985).
7. Porter, J. W. et al. *Proc. Natl Acad. Sci. USA* **86**, 9342–9346 (1989).
8. McField, M. D. *Bull. Mar. Sci.* **64**, 155–172 (1999).
9. Mumby, P. J. *Mar. Ecol. Prog. Ser.* **190**, 27–35 (1999).
10. Aronson, R. B. & Precht, W. F. *Paleobiology* **23**, 326–346 (1997).
11. Macintyre, I. G. et al. *Atoll Res. Bull.* **430**, 1–18 (1995).
12. Smith, S. V. & Buddemeier, R. W. *Annu. Rev. Ecol. Syst.* **23**, 89–118 (1992).
13. Kleypas, J. A. et al. *Science* **284**, 118–120 (1999).
14. Harvell, C. D. et al. *Science* **285**, 1505–1510 (1999).

Genome stability

Transgenerational mutation by radiation

Parental exposure to ionizing radiation increases the frequency of germline mutations detectable in the next generation¹. Parental exposure can also increase the rate of mutation in somatic cells^{2,3} and confer a predisposition to cancer^{4–6} in offspring, suggesting that there could be an indirect effect of radiation on somatic genome stability that is transmissible through the germ line of the irradiated parents. We have found that this indirect effect extends to the germ line of unexposed first-generation offspring in mice, as revealed by an increased instability of repeat-DNA sequences in their descendants.

CBA/H male mice were exposed to fission neutrons and mated to unexposed females (Fig. 1a). Exposure resulted in a 6-fold increase in paternal mutation rate at expanded simple-tandem repeat (ESTR) loci^{7,8} detectable in first-generation (F₁) offspring (Table 1). Breeding from these unexposed F₁ mice revealed that germline mutation rates remained significantly high in transmissions from both F₁ males and F₁ females (6-fold and 3.5-fold increase, respectively, compared to unexposed control families and to the alleles transmitted from control breeding partners; not shown).

This increase was in part due to increased mutational mosaicism in the germ line of the F₁ mice, with more than 50% of *de novo* mutations detected in F₂ mice being shared by two or more littermates, compared with less than 25% in the offspring of both non-irradiated mice and F₀ mice exposed to X-rays⁹ or fission neutrons (Table 1 and Fig. 1b). However, singleton mutations (those present in only one member of a litter) were also more frequent in these F₂ mice than in the controls, significantly so for male transmissions (Table 1). We conclude that paternal exposure to radiation results in the destabilization of ESTR loci in the germ line of offspring, and that some of the resulting mutations occur sufficiently early in germline development for significant levels of mosaicism to arise.

This remarkable stimulation of muta-

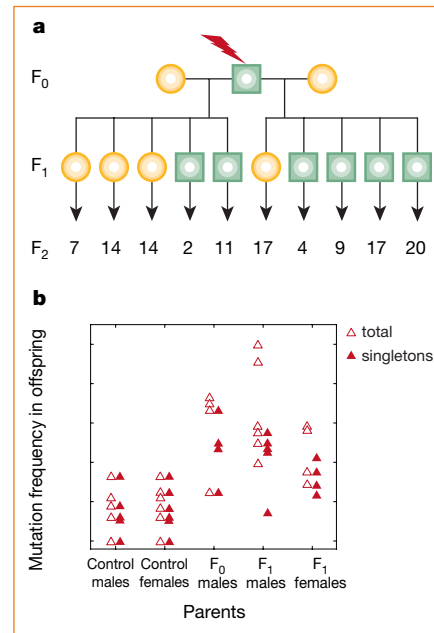


Figure 1 Expanded simple-tandem repeat mutation in the descendants of irradiated male mice. **a**, Design of the transgenerational study: the number of F₂ offspring derived from each F₁ mouse is indicated. **b**, Frequency of ESTR mutation in offspring of F₀ and F₁ males and females. Four CBA/H males were given whole-body chronic irradiation of 0.5 Gy of fission neutrons (²⁵²Cf source, 0.03 Gy min⁻¹) and 10 weeks later mated to non-irradiated CBA/H females, ensuring that litters were derived from irradiated stem cells. Litters from one irradiated male were used for subsequent breeding of F₂ mice. DNA profiles were produced⁹ using the mouse single-locus ESTR probes *Ms6-hm* and *Hm-2* (refs 7,8).

tion in the F₁ germ line following paternal irradiation is reminiscent of delayed radiation-induced genome instability in somatic cells, where radiation can induce genetic changes in some of the descendants of a single irradiated cell, an effect that can persist for many cell generations¹⁰. By analogy, exposure of F₀ mice must create a signal in the F₀ germ line that is transmitted by a single sperm to the F₁ offspring, in which instability arises many cell generations later and is detected in some but not all cells in the developing germ line.

It is unlikely that the negligible cytoplasmic component of the mature sperm could carry substantial amounts of long-lived free radicals or other radiation-induced species into the egg, so the transmitted signal is probably DNA-dependent. It is most

unlikely that this signal is radiation-induced damage at the ESTR loci themselves, because it is transmissible through meiosis and mitosis and appears to operate *in trans* in the F₁ germ line, affecting alleles both from the exposed F₀ male and from the unexposed F₀ female (data not shown).

Increased ESTR mutation rates appear to be uniform in the germ line of all F₁ offspring derived from 10 independent sperm from the irradiated F₀ male (Fig. 1b; χ^2 test for homogeneity, $P > 0.5$), suggesting that radiation-induced *de novo* mutations in DNA-repair genes are not the cause of instability in the F₁ germ line. The alternative, therefore, is a radiation-exposure signal inherited through sperm in an epigenetic fashion, perhaps through changes in DNA methylation. Methylation is transmissible through many cell divisions¹¹ and can influence the activity of DNA-repair systems¹². If radiation-induced DNA damage in the germ line triggers an epigenetic signal that influences DNA repair and can survive the reprogramming of DNA methylation during spermatogenesis and early development, then subsequent activation of this signal in the germ line could lead to the observed transgenerational mutagenesis.

These findings have potential implications for risk evaluation in humans. So far, the main concern of radiation protection is focused on the direct mutagenic effects of ionizing radiation in exposed individuals¹. However, the persistence of instability into the germ line of unexposed offspring of irradiated mice raises the issue of delayed genetic risk, as well as the potential for radiation to cause mutation clustering and to contribute to mosaicism in germ cells, long recognized as an important mechanism in the origin of human genetic disorders¹³.

Yuri E. Dubrova*†, Mark Plumb‡, Bruno Gutierrez*, Emma Boulton‡, Alec J. Jeffreys*

*Department of Genetics, University of Leicester, Leicester LE1 7RH, UK
e-mail: yed2@le.ac.uk

†N. I. Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow B-333, Russia

‡Medical Research Council Radiation and Genome Stability Unit, Harwell, Oxon OX11 0RD, UK

1. UNSCEAR Sources and Effects of Ionizing Radiation (United Nations, New York, 1993).
2. Luke, G. A., Riches, A. C. & Bryant, P. E. *Mutagenesis* **12**, 147–152 (1997).
3. Burrue, V. R., Raabe, O. G. & Wiley, L. M. *Mutat. Res.* **381**, 59–66 (1997).
4. Nomura, T. *Mutat. Res.* **121**, 59–68 (1983).
5. Vorobtsova, I. A., Aliyapova, C. M. & Anisimov, V. N. *Mutat. Res.* **287**, 207–216 (1993).
6. Lord, B. I. *et al. Br. J. Cancer* **78**, 301–311 (1998).
7. Kelly, R. *et al. Genomics* **5**, 844–856 (1989).
8. Gibbs, M. *et al. Genomics* **17**, 121–128 (1993).
9. Dubrova, Y. E. *et al. Proc. Natl Acad. Sci. USA* **95**, 6251–6255 (1998).
10. Morgan, W. F. *et al. Radiat. Res.* **146**, 247–258 (1996).
11. Holliday, R. *Science* **238**, 163–170 (1987).
12. Herman, J. G. *et al. Proc. Natl Acad. Sci. USA* **95**, 6870–6875 (1998).
13. Zlatogora, J. *Hum. Genet.* **102**, 381–386 (1988).

Table 1 Mutation in two mouse generations at ESTR loci *Ms6-hm* and *Hm-2*

Group	Offspring (litters)	All mutations				Singletons			
		Mutations	Rate	Ratio*	P†	Mutations‡	Rate	Ratio*	P†
Control males	74 (12)	8	0.0540	-	-	6 (75%)	0.0405	-	-
Control females	74 (12)	9	0.0608	-	-	7 (77.8%)	0.0473	-	-
F ₀ males, 0.5 Gy of neutrons	34 (7)	22	0.3235	5.98	8.62 × 10⁻⁷	16 (72.7%)	0.2353	5.80	6.59 × 10⁻⁵
F ₁ males	63 (11)	38	0.3016	5.58	5.81 × 10⁻⁸	14 (36.8%)	0.1111	2.74	0.0444
F ₁ females	52 (6)	22	0.2115	3.45	0.0007	11 (50%)	0.1058	2.24	0.1291

*Ratio to control.

†Probability of difference from the control group (Fisher's exact test, two-tailed; statistically significant values are in bold).

‡Percentage of singleton mutations is given in parentheses.

Materials

Spontaneous formation of inorganic helices

Attempts to generate films, wires, helices, rings and regular patterns out of inorganic materials^{1–10} have been going on for the past 100 years. We show here that stable helices of porous manganese oxide materials can be formed spontaneously from uniform sols and that they are excellent semiconductors. These inorganic helices contain micropores, can be converted into other structures, and their composition can be varied.

Tetramethyl ammonium (TMA⁺) permanganate salts in 2-butanol/H₂O form sols that when heated in open capillary tubes can form ring and helical gels (Table 1 and Fig. 1). Helices are produced from concentrated sols and separated rings at low concentrations (10^{–3}). The spacing between rings increases from the top to the bottom end of the capillary, as in the spacing law demonstrated by Liesegang rings¹¹.

When helices form, the number of turns per unit length is primarily related to the starting concentration of the sol, whereas their diameters are mainly influenced by the capillary diameter and range from millimetres to less than 30 µm, with lengths of up to 25 cm. Free-standing helices can be obtained by cutting the tubes, or by using water to detach them from the capillary wall.

The formation of helices starts first by virtue of a contraction of the sol–gel as a result of solvent evaporation. The elastic gel then buckles into a helical shape, possibly because of the capillary pressure applied at its upper end by the surrounding fluid. Further evaporation of solvent eliminates

this capillary pressure and the gel expands from the bottom to the top of the capillary (see <http://www.lib.uconn.edu/chemistry/SuibGroup/suibg.html>). We are investigating the mechanism of this process by quantifying the contribution and interplay of gel anisotropy and elasticity, wetting and dewetting, capillarity, and agglomeration of the sol particles that combine to produce these remarkable shapes.

Ion exchange of the TMA helix with K⁺ leads to crystalline helices with composition K_{0.93}Mn_{2.1}⁴⁺Mn_{1.9}³⁺O₇(OH)_{1.03}·2.7H₂O; the lattice parameters are similar to those of synthetic birnessite (OL-1; ref. 12). Ion exchange is rapid, indicating that these materials are highly permeable. Thermal treatment of the helices of K-OL-1 at 500 °C leads to the formation of an octahedral molecular-sieve (OMS) tunnel structure of synthetic cryptomelane (OMS-2), and a composition of K_{1.43}⁺Mn_{1.86}⁴⁺Mn_{4.7}³⁺O_{14.57}(OH)_{1.43}·0.7H₂O with an average oxidation state for manganese of 3.58.

We examined the K-OMS-2 helices by scanning and transmission electron microscopy. Striations are evident along the length, indicative of a fibrous structure, corroborated by the direction of crack propagation from sample preparation. The helix cores are usually solid, however, with secondary spiralling along the radial direction of the helix being sometimes seen. The helix cores are usually solid.

The microstructure leads to microporosity of the 2 × 2 crystal structure of OMS-2, mesoporosity in between the colloidal crystallites, and a net alignment and connectivity of the microporous network along the length of the helix. This alignment leads to conduction along the helix and, as a result of the porosity, rate processes such as ion exchange are enhanced owing to fast dif-

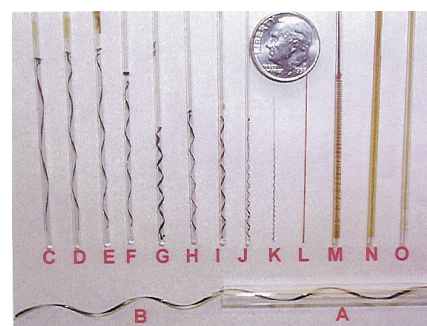


Figure 1 Morphology of self-assembled (TMA), MnO, structures. Different samples A–O represent colloids produced by varying manganese concentration, capillary diameter, temperature, and heating period (Table 1). The encapsulated self-assembled products are shown as the final morphology attained after solvent evaporation at 85 °C. Coin diameter is 18 mm.

fusion. The anisotropic (that is, larger by an order of magnitude in the direction parallel to the helix axis compared with the perpendicular direction) DC-4 probe conductivity of $4.2 \times 10^{-1} \text{ W}^{-1} \text{ cm}^{-1}$ at 21 °C for the K-OMS-2 system is roughly an order of magnitude more conductive than most well-formed single crystals of semiconducting cryptomelane-like materials¹².

Oscar Giraldo*, Stephanie L. Brock*, Manuel Marquez*, Steven L. Suib*†‡, Hugh Hillhouse§, Michael Tsapatsis§

Departments of *Chemistry and †Chemical Engineering and ‡Institute of Materials Science, University of Connecticut, Storrs, Connecticut 06269-3060, USA

e-mail: suib@uconnvm.uconn.edu

§Department of Chemical Engineering, 159 Goessmann Laboratory, University of Massachusetts, Amherst, Massachusetts 01003-3110, USA

- Wu, C. G. & Bein, T. *Science* **264**, 1757–1759 (1994).
- Trau, M. et al. *Nature* **390**, 674–676 (1997).
- den Exter, M. J. et al. in *Recent Advances and New Horizons in Zeolite Science and Technology* (ed. Chon, H.) Vol. 102, 413–454 (Stud. Surf. Sci. Catal., Elsevier, Amsterdam, 1996).
- Jones, C. W., Tsuji, K. & Davis, M. E. *Nature* **393**, 52–54 (1998).
- Davis, W. B. et al. *Nature* **396**, 60–63 (1998).
- Rowan, A. E. & Nolte, R. J. M. *Angew. Chem. Int. Ed.* **37**, 63–68 (1998).
- Soghomonian, V. et al. *Science* **259**, 1596–1599 (1993).
- Gier, T. E., Bu, X., Feng, P. & Stucky, G. D. *Nature* **395**, 154–157 (1998).
- Raimondi, M. E., Maschmeyer, T., Templer, R. H. & Seddon, J. M. *J. Chem. Soc. Chem. Comm.* **19**, 1843–1844 (1997).
- Bruinsma, P. J. et al. *Chem. Mater.* **9**, 2507–2512 (1997).
- Antal, T., Droz, M., Magnin, J. & Racz, Z. *Phys. Rev. Lett.* **83**, 2880–2883 (1999).
- Shen, Y. F. et al. *Science* **260**, 511–515 (1993).

Erratum

Stomach rinsing in rays

D. W. Sims, P. L. R. Andrews, J. Z. Young
Nature **404**, 566 (2000)

An editing error caused a misrepresentation of results reported in ref. 8: in the first sentence of the third-to-last paragraph, the intended meaning is that cloacal protrusion has been described in the Carcharhinidae, and not peritoneal-inflation-induced cloacal protrusion as the text implies.

Table 1 Conditions leading to formation of rings and helices

Sample	ID (mm)	Time (h)	[Mn] (M)	Observations
A	3.0	140	1.0×10^{-1}	Formation inside the glass tube
B	3.0	140	1.0×10^{-1}	Formation on the external surface of the glass tube
C	1.2	24	1.0×10^{-1}	Formation inside the glass tube
D	1.2	24	9.0×10^{-2}	Formation inside the glass tube
E	1.2	24	8.0×10^{-2}	Formation inside the glass tube
F	1.2	24	7.0×10^{-2}	Formation inside the glass tube
G	1.2	24	5.0×10^{-2}	Formation inside the glass tube
H	1.2	48	5.0×10^{-2}	Formation inside the glass tube
I	1.2	24	5.0×10^{-2}	Helix post-heated at 500 °C for 2 h
J	0.8	24	1.0×10^{-1}	Helix formed in a quartz tube
K	0.5	24	1.0×10^{-1}	Helix formed in a glass tube
L	0.2	24	1.0×10^{-1}	Helix formed in a GC capillary tube
M	1.2	24	5.0×10^{-3}	Ring formation
N	1.2	24	1.0×10^{-3}	Mixture of rings and helices formed
O	1.2	24	1.0×10^{-3}	Only micro-rings formed

ID, internal diameter of capillary tube; [Mn], manganese concentration (molar). A colloidal solution of lamellar manganese oxide (0.1 M in Mn) was prepared by stirring 10 mol [(CH₃)₄N]⁺MnO₄[–] in 100 ml of distilled deionized water and 30 ml of 2-butanol at room temperature. After 30 min, a dark-red–brown solution formed in the lower aqueous layer was separated off and used undiluted or diluted (down to 0.001 M in Mn). Solutions were put into open-ended capillaries. Angles of tubes between 45° and 90° give well-formed helices. Solvent evaporation in an oven at 85 °C produced helices as the sol shrank to form a gel (< 12 h). A gel of [(CH₃)₄N]⁺Mn_{2.1}⁴⁺Mn_{1.9}³⁺O₇(OH)_{1.03}·5H₂O, with a manganese average oxidation state of 3.52, was formed after heating at 85 °C for 160 h in capillary tubes of ID 3 mm. Theoretical composition (experimental via inductively coupled plasma and combustion analyses in parentheses): for TMA, MnO, helix, %C 8.82 (10.58), %H 4.36 (4.47), %N 2.57 (2.86); for K-OL-1 helix, %C 0 (0.34), %H 1.47 (1.23), %N 0 (0.03), K/Mn 0.2325 (0.2325); for K-OMS-2 helix, K/Mn 0.2325 (0.2325). The flexibility of the helices may be due to elastic anisotropy resulting from the fibrous texture of the gel.

A surprising simplicity to protein folding

David Baker

Department of Biochemistry, University of Washington, J567 Health Sciences Building, Box 357350, Seattle, Washington 98195, USA

The polypeptide chains that make up proteins have thousands of atoms and hence millions of possible inter-atomic interactions. It might be supposed that the resulting complexity would make prediction of protein structure and protein-folding mechanisms nearly impossible. But the fundamental physics underlying folding may be much simpler than this complexity would lead us to expect: folding rates and mechanisms appear to be largely determined by the topology of the native (folded) state, and new methods have shown great promise in predicting protein-folding mechanisms and the three-dimensional structures of proteins.

Proteins are linear chains of amino acids that adopt unique three-dimensional structures ('native states') which allow them to carry out intricate biological functions. All of the information needed to specify a protein's three-dimensional structure is contained within its amino-acid sequence. Given suitable conditions, most small proteins will spontaneously fold to their native states¹.

The protein-folding problem can be stated quite simply: how do amino-acid sequences specify proteins' three-dimensional structures? The problem has considerable intrinsic scientific interest: the spontaneous self-assembly of protein molecules with huge numbers of degrees of freedom into a unique three-dimensional structure that carries out a biological function is perhaps the simplest case of biological self-organization. The problem also has great practical importance in this era of genomic sequencing: interpretation of the vast amount of DNA sequence information generated by large-scale sequencing projects will require determination of the structures and functions of the encoded proteins, and an accurate method for protein structure prediction could clearly be vital in this process.

Since Anfinsen's original demonstration of spontaneous protein refolding, experimental studies have provided much information on the folding of natural proteins²⁻⁴. Complementary analytical and computational studies of simple models of folding have provided valuable and general insights into the folding of polymers and the properties of folding free-energy landscapes⁵⁻⁷. These studies of

idealized representations of proteins have inspired new models, some described here, which attempt to predict the results of experimental measurements on real proteins.

Because the number of conformations accessible to a polypeptide chain grows exponentially with chain length, the logical starting point for the development of models attempting to describe the folding of real protein is experimental data on very small proteins (fewer than 100 residues). Fortunately, there has been an explosion of information about the folding of such small proteins over the last ten years³. For most of these proteins, partially ordered non-native conformations are not typically observed in experiments, and the folding reactions can usually be well modelled as a two-state transition between a disordered denatured state and the ordered native state. In contrast, the folding kinetics of larger proteins may in some cases be dominated by escape from low-free-energy non-native conformations. The folding of larger proteins is also often facilitated by 'molecular chaperones'⁸ which prevent improper protein aggregation.

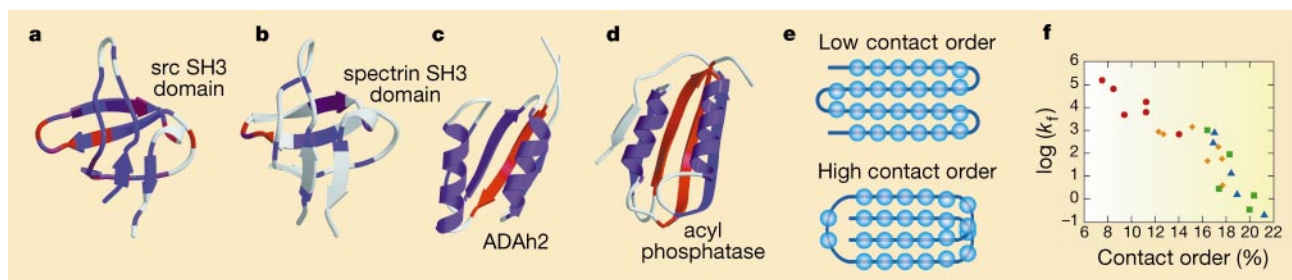
To pass between the unfolded and native low-free-energy states, the protein must pass through a higher-free-energy transition state. In the unfolded state the protein can take up any one of many conformations, whereas in the native state it has only one or a few distinct conformations. The degree of heterogeneity of conformations in the transition state has thus been the subject of much discussion⁹⁻¹¹. For example, one of the main differences between the

Box 1

Dependence of folding mechanisms on topology

The structures of folding transition states are similar in proteins with similar native structures. The distribution of structure in the transition state ensemble can be probed by mutations at different sites in the chain; mutations in regions that make stabilizing interactions in the transition state ensemble slow the folding rate, whereas mutations in regions that are disordered in the transition state ensemble have little effect⁴. For example, in the structures of the SH3 domains of src¹⁸ (**a**) and spectrin¹⁷ (**b**), and the structurally related proteins Adah2 (ref. 37; **c**) and acyl phosphatase¹⁶ (**d**), the colours code for the effects of mutations on the folding rate. Red, large effect; magenta, moderate effect; and blue, little effect. In the two SH3 domains, the turn coloured in red at the left of the structures appears to be largely formed, and the beginning and end of the protein largely disrupted, in the transition state ensemble. (To facilitate

the comparison in **c** and **d**, the average effect of the mutations in each secondary structure element is shown.) This dependence of folding rate on topology has been quantified by comparing folding rates and the relative contact order of the native structures. The relative contact order is the average separation along the sequence of residues in physical contact in a folded protein, divided by the length of the protein. **e**, A low- and high-contact-order structure for a four-strand sheet. In **f**, black circles represent all-helical proteins, green squares sheet proteins and red diamonds proteins comprising both helix and sheet structures. The correlation between contact order and folding rate (k_f) is striking, occurring both within each structural subclass and within sets of proteins with similar overall folds (proteins structurally similar to the α/β protein acyl phosphatase¹⁶ are indicated by blue triangles).



'old' and 'new' views of protein folding is that the 'new' view allows for a much more heterogeneous transition state—really a transition state ensemble—than the 'old' view, which concentrated on a single, well defined folding 'pathway'.

The primary measurements that can be made experimentally of the highly cooperative folding reactions of small proteins are: the folding rate; the distribution of structures in the transition state ensemble, inferred from the effects of mutations on the folding rate (Box 1); and the structure of the native state. Here I focus on recent progress in predicting these three features.

Topology determines folding mechanisms

Are simple models likely to be able to account for the overall features of the folding process, given the many possible inter-atomic interactions in even a small protein? Recent data indicate that the fundamental physics underlying the folding process may be simpler than was previously thought.

The complexity of protein structure emerges from the details of how individual atoms in both a protein's peptide backbone and its

amino-acid residues interact. However, the general path that the polymer chain takes through space—its topology—can be very similar between proteins. Three independent lines of investigation indicate that protein-folding rates and mechanisms are largely determined by a protein's topology rather than its inter-atomic interactions¹².

First, large changes in amino-acid sequence, either experimental^{13,14} or evolutionary¹⁵, that do not alter the overall topology of a protein usually have less than tenfold effect on the rate of protein folding¹⁵. This suggests evolution has not optimized protein sequences for rapid folding, an encouraging result for simple model development.

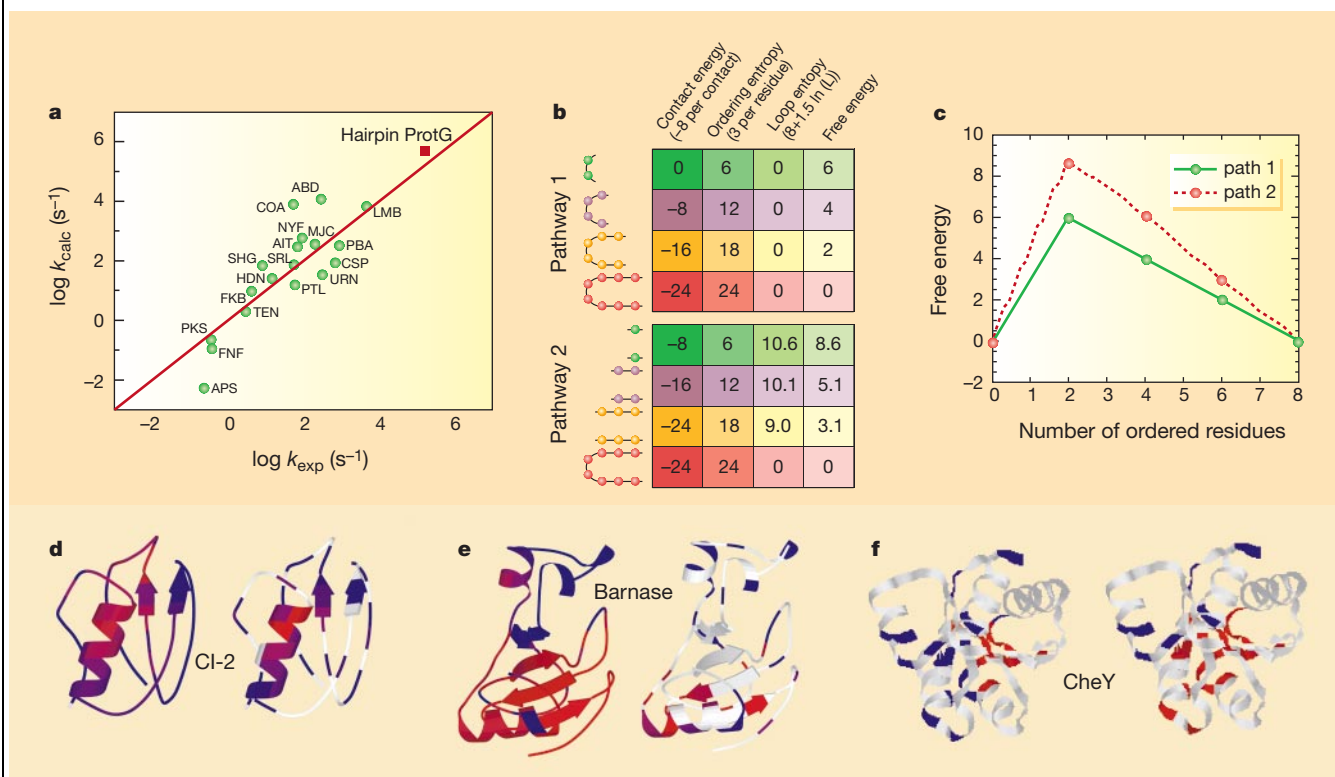
Second, using the consequences of mutations on folding kinetics to probe the transition states of proteins with similar structures but very different sequences has shown that the structures of these transition states are relatively insensitive to large-scale changes in sequence^{16–18}. For example, in Box 1 there are two examples of pairs of structurally related proteins with little or no sequence similarity that have very similar folding transition-state ensembles.

Box 2

Prediction of protein-folding mechanisms

Munoz and Eaton²⁴ computed folding rates by solving the diffusion equation of motion on the one-dimensional free-energy profiles that result from projection of the full free-energy landscape onto a reaction coordinate corresponding to the number of ordered residues. **a** shows the accuracy of their prediction by plotting computed folding rates (k_{calc}) against experimentally measured rates (k_{exp}). To predict folding transition state structure, the lowest free energy paths to the native state can be identified. For example, a β -hairpin (**b**) has two possible paths to the native state, beginning at the hairpin (pathway 1) or at the free ends (pathway 2; ordered residues only are indicated; L is loop length). The Table gives the contributions to the free energy of each configuration (total free energy is the sum of the first three columns). Plotting the free energy as a function of the number of ordered residues (**c**) shows that the transition state for both pathways consists of configurations with two of the residues ordered. Calculations on real proteins (**d–f**) have considered

all possible paths: the folding rate and transition state structure are determined from the lowest free-energy paths. Galzitskaya and Finkelstein²⁵ and Alm and Baker²⁶ predicted the folding transition state structure of CheY (**f**), and Cl-2 (**d**) and barnase (**e**), respectively. They identified the transition-state ensemble by searching for the highest free-energy configurations on the lowest free-energy paths between unfolded and folded states. The effects of mutations on the folding rate were predicted on the basis of the contribution of the interactions removed by the mutations to the free energy of the transition state ensemble, or by directly determining the change in folding rate. The predicted effects of mutations on the folding rates are shown on the native structure (left); the measured effects, on the right (the colour scheme is as in Box 1; grey, regions not probed by mutations; experimental results for Cl-2 and barnase, ref. 4; CheY, ref. 38).



Third, the folding rates of small proteins correlate with a property of the native state topology: the average sequence separation between residues that make contacts in the three-dimensional structure (the 'contact order'; Box 1). Proteins with a large fraction of their contacts between residues close in sequence ('low' contact order) tend to fold faster than proteins with more non-local contacts ('high' contact order)^{12,19}. This correlation holds over a million-fold range of folding rates, and is remarkable given the large differences in the sequences and structures of the proteins compared. Simple geometrical considerations appear to explain much of the difference in the folding rates of different proteins.

The important role of native-state topology can be understood by considering the relatively large entropic cost of forming non-local interactions early in folding. The formation of contacts between residues that are distant along the sequence is entropically costly, because it greatly restricts the number of conformations available to the intervening segment. Thus, interactions between residues close together in sequence are less disfavoured early in folding than interactions between widely separated residues. So, for a given topology, local interactions are more likely to form early in folding than non-local interactions. Likewise, simple topologies with mostly local interactions are more rapidly formed than those with many non-local interactions. More generally, the amount of configurational entropy lost before substantial numbers of favourable native interactions can be made depends on the topology of the native state. The importance of topology has also been noted in studies of computational models of folding^{20–23}.

As proteins' sequences determine their three-dimensional structures, both protein stability and protein-folding mechanisms are ultimately determined by the amino-acid sequence. But whereas stability is sensitive to the details of the inter-atomic interactions (removal of several buried carbon atoms can completely destabilize a protein), folding mechanisms appear to depend more on the low-resolution geometrical properties of the native state.

Predicting folding mechanism from topology

The results described above indicate that simple models based on the structure of the native state should be able to predict the coarse-grained features of protein-folding reactions. Several such models have recently been developed, and show considerable promise for predicting folding rates and folding transition-state structures. Three approaches^{24–26} have attempted to model the trade-off between the formation of attractive native interactions and the loss of configurational entropy during folding. Each assumes that the only favourable interactions possible are those formed in the native state. This neglect of non-native interactions is consistent with the observed importance of native-state topology in folding, and dates back to the work of Go on simple lattice models²⁷.

Although the approaches differ in detail, the fundamental ideas are similar. All use a binary representation of the polypeptide chain in which each residue is either fully ordered, as in the native state, or completely disordered. To limit the number of possible configurations, all ordered residues are required to form a small number of segments, continuous in sequence. Attractive interactions are taken to be proportional to the number of contacts, or the amount of buried surface area, between the ordered residues in the native structure, and non-native interactions are completely ignored. The entropic cost of ordering is a function of the number of residues ordered and the length of the loops between the ordered segments. Folding kinetics are modelled by allowing only one residue to become ordered (or disordered) at a time. As the number of ordered residues increases, the free energy first increases, owing to the entropic cost of chain ordering, and then decreases, as large numbers of attractive native interactions are formed.

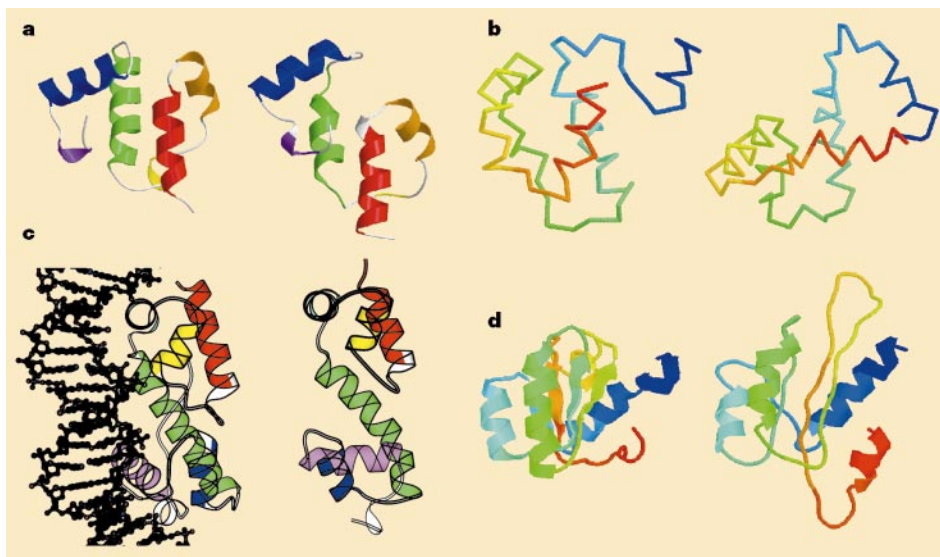
Such simple models can potentially be used to predict experimentally measurable quantities such as the folding rate, which depends on the height of the free-energy barrier, and the effects of mutations on the folding rate, which depend on the region(s) of the protein ordered near the top of the barrier. Predictions of both

Box 3

Ab initio structure predictions

Blind *ab initio* structure predictions for the CASP3 protein structure prediction experiment. For each target, the native structure is shown on the left with a good prediction on the right (predictions by Baker³⁹(**a**, **c**), Levitt⁴⁰(**b**) and Skolnick⁴¹(**d**) and colleagues; for more information see <http://predictioncentre.llnl.gov/> and *Proteins* Suppl. **3**, 1999). Segments are colour coded according to their position in the sequence (from blue (amino terminus) to red (carboxy terminus)). **a**, DNA B helicase⁴¹. This protein had a novel fold and thus could not be predicted using standard fold-recognition methods. Not shown are N- and

C-terminal helices which were positioned incorrectly in the predicted structure. **b**, Ets-1 (ref. 43). **c**, MarA⁴⁴. This prediction had potential for functional insights; the predicted two-lobed structure suggests the mechanism of DNA binding (left, X-ray structure of the protein–DNA complex). **d**, L30. A large portion of this structure was similar to a protein in the protein databank but the best *ab initio* predictions were competitive with those using fold-recognition methods. The three approaches that produced these predictions used reduced-complexity models for all or almost all of the conformational search process.



folding rates and folding transition-state structures using these simple models are quite encouraging (Box 2; other recent models have also yielded good results^{28–33}).

The success of these models in reproducing features of real folding reactions again supports the idea that the topology of the native state largely determines the overall features of protein-folding reactions and that non-native interactions have a relatively minor role. Incorporation of sequence-specific information into these models, either in the inter-residue interactions or in the free-energy costs of ordering different segments of the chain, should improve their accuracy to the point where they may be able to account for much of the experimental data on the folding of small proteins.

Ab initio structure prediction

Predicting three-dimensional protein structures from amino-acid sequences alone is a long-standing challenge in computational molecular biology. Although the preceding sections suggest that the only significant basin of attraction on the folding landscapes of small proteins is the native state, the potentials used in *ab initio* structure-prediction efforts have not had this property, and until recently such efforts met with little success. The results of an international blind test of structure prediction methods (CASP3; ref. 34) indicate, however, that significant progress has been made^{35,36}.

As with the models for protein-folding mechanisms, most of the successful methods attempt to ignore the complex details of the inter-atomic interactions—the amino-acid side chains are usually not explicitly represented—and instead focus on the coarse-grained features of sequence–structure relationships. Problems in which the full atomic detail of interactions in the native state is important—such as the design of novel stable proteins, and the prediction of stability and high resolution structure—will almost certainly require considerably more detailed models.

Some of the most successful blind *ab initio* structure predictions made in CASP3 are shown in Box 3. In several of these predictions the root-mean-square deviation between backbone carbon atoms in the predicted and experimental structures is below 4.0 Å over segments of up to 70 residues. Several of these models can compete with more traditional fold-recognition methods. At least one case (Mar A) gave a model capable of providing clues about protein function³⁹.

The predictions are an encouraging improvement over those achieved in the previous structure-prediction experiment (CASP2), but improvements are still needed to the accuracy and reliability of the models. Improvements in *ab initio* structure prediction may allow these methods to generate reliable low-resolution models of all the small globular proteins in an organism's genome.

Emerging simplicity

The experimental results and predictions discussed here indicate that the fundamental physics underlying folding may be simpler than previously thought and that the folding process is surprisingly robust. The topology of a protein's native state appears to determine the major features of its folding free-energy landscape. Both protein structures and protein-folding mechanisms can be predicted, to some extent, using models based on simplified representations of the polypeptide chain. The challenge ahead is to improve these models to the point where they can contribute to the interpretation of genome sequence information. □

1. Anfinsen, C. Principles that govern the folding of protein chains. *Science* **181**, 223–227 (1973).
2. Baldwin, R. L. & Rose, G. D. Is protein folding hierarchic? II. Folding intermediates and transition states. *Trends Biochem. Sci.* **24**, 26–33 (1999).
3. Jackson, S. E. How do small single-domain proteins fold? *Fold. Des.* **3**, R81–91 (1998).
4. Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (Freeman, New York, 1999).

5. Chan, H. S. & Dill, K. A. Protein folding in the landscape perspective: chevron plots and non-Arrhenius kinetics. *Proteins* **30**, 2–33 (1998).
6. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167–195 (1995).
7. Dobson, C. M. & Karplus, M. The fundamentals of protein folding: bringing together theory and experiment. *Curr. Opin. Struct. Biol.* **9**, 92–101 (1999).
8. Horwich, A. L. Chaperone rings in protein folding and degradation. *Proc. Natl Acad. Sci. USA* **96**, 11033–11040 (1999).
9. Shakhnovich, E. I. Folding nucleus: specific or multiple? Insights from lattice models and experiments. *Fold. Des.* **3**, R108–111 (1998).
10. Pande, V. S., Grosberg, A. Y., Tanaka, T. & Rokhsar, D. Pathways for protein folding: is a new view needed? *Curr. Opin. Struct. Biol.* **8**, 68–79 (1998).
11. Thirumalai, D. & Klimov, D. K. Fishing for folding nuclei in lattice models and proteins. *Fold. Des.* **3**, R112–118 (1998).
12. Alm, E. & Baker, D. Matching theory and experiment in protein folding. *Curr. Opin. Struct. Biol.* **9**, 189–196 (1999).
13. Riddle, D. S. *et al.* Functional rapidly folding proteins from simplified amino acid sequences. *Nature Struct. Biol.* **4**, 805–809 (1997).
14. Kim, D. E., Gu, H. & Baker, D. The sequences of small proteins are not extensively optimized for rapid folding by natural selection. *Proc. Natl Acad. Sci. USA* **95**, 4982–4986 (1998).
15. Perl, D. *et al.* Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. *Nature Struct. Biol.* **5**, 229–235 (1998).
16. Chiti, F. *et al.* Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nature Struct. Biol.* **6**, 1005–1009 (1999).
17. Martinez, J. C. & Serrano, L. The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nature Struct. Biol.* **6**, 1010–1016 (1999).
18. Riddle, D. S. *et al.* Experiment and theory highlight role of native state topology in SH3 folding. *Nature Struct. Biol.* **6**, 1016–1024 (1999).
19. Plaxco, K. W., Simons, K. T. & Baker, D. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985–994 (1998).
20. Shea, J. E., Onuchic, J. N. & Brooks, C. L. III. Exploring the origins of topological frustration: design of a minimally frustrated model of fragment B of protein A. *Proc. Natl Acad. Sci. USA* **96**, 12512–12517 (1999).
21. Onuch, J. N., Nymeyer, H., Garcia, A. E., Chaine, J. & Socci, N. D. The energy landscape theory of protein folding: insights into folding mechanism and scenarios. *Adv. Protein Chem.* **53**, 87–152 (2000).
22. Micheletti, C., Banavar, J. R., Maritan, A. & Seno, F. Protein structures and optimal folding from a geometrical variational principle. *Phys. Rev. Lett.* **82**, 3372–3375 (1999).
23. Abkevich, V., Gutin, A. & Shakhnovich, E. Specific nucleus as the transition state for protein folding: evidence from the lattice model. *Biochemistry* **33**, 10026–10036 (1994).
24. Munoz, V. & Eaton, W. A. A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc. Natl Acad. Sci. USA* **96**, 11311–11316 (1999).
25. Galzitskaya, O. V. & Finkelstein, A. V. A theoretical search for folding/unfolding nuclei in three-dimensional protein structures. *Proc. Natl Acad. Sci. USA* **96**, 11299–11304 (1999).
26. Alm, E. & Baker, D. Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proc. Natl Acad. Sci. USA* **96**, 11305–11310 (1999).
27. Go, N. Theoretical studies of protein folding. *Annu. Rev. Biophys. Bioeng.* **12**, 183–210 (1983).
28. Portman, J. J., Takada, S. & Wolynes, P. G. Variational theory for site resolved protein folding free energy surfaces. *Phys. Rev. Lett.* **81**, 5237–5240 (1998).
29. Debye, D. & Goddard, W. A. First principles prediction of protein-folding rates. *J. Mol. Biol.* **294**, 619–625 (1999).
30. Li, A. J. & Daggett, V. Identification and characterization of the unfolding transition state of chymotrypsin inhibitor 2 by molecular dynamics simulations. *J. Mol. Biol.* **257**, 412–429 (1996).
31. Lazaridis, T. & Karplus, M. 'New view' of protein folding reconciled with the old through multiple unfolding simulations. *Science* **278**, 1928–1931 (1997).
32. Sheinerman, F. B. & Brooks, C. L. III. A molecular dynamics simulation study of segment B1 of protein G. *Proteins* **29**, 193–202 (1997).
33. Burton, R. E., Myers, J. K. & Oas, T. G. Protein folding dynamics: quantitative comparison between theory and experiment. *Biochemistry* **37**, 5337–5343 (1998).
34. Moul, J., Hubbard, T., Fidelis, K. & Pedersen, J. T. Critical assessment of methods of protein structure prediction (CASP): round III. *Proteins (Suppl.)* **3**, 2–6 (1999).
35. Orengo, C. A., Bray, J. E., Hubbard, T., LoConte, L. & Sillitoe, I. Analysis and assessment of *ab initio* three-dimensional prediction, secondary structure, and contacts prediction. *Proteins (Suppl.)* **3**, 149–170 (1999).
36. Murzin, A. G. Structure classification-based assessment of CASP3 predictions for the fold recognition targets. *Proteins (Suppl.)* **3**, 88–103 (1999).
37. Villegas, V., Martinez, J. C., Aviles, F. X. & Serrano, L. Structure of the transition state in the folding process of human procarboxypeptidase A2 activation domain. *J. Mol. Biol.* **283**, 1027–1036 (1998).
38. Lopez-Hernandez, E. & Serrano, L. Structure of the transition state for folding of the 129 aa protein CheY resembles that of a smaller protein, CI-2. *Fold. Des.* **1**, 43–55 (1996).
39. Simons, K. T., Bonneau, R., Ruczinski, I. & Baker, D. *Ab initio* protein structure prediction of CASP III targets using ROSETTA. *Proteins (Suppl.)* **3**, 171–176 (1999).
40. Samudrala, R., Xia, Y., Huang, E. & Levitt, M. *Ab initio* protein structure prediction using a combined hierarchical approach. *Proteins (Suppl.)* **3**, 194–198 (1999).
41. Ortiz, A. R., Kolinski, A., Rotkiewicz, P., Ilkowsky, B. & Skolnick, J. *Ab initio* folding of proteins using restraints derived from evolutionary information. *Proteins (Suppl.)* **3**, 177–185 (1999).
42. Weigelt, J., Brown, S. E., Miles, C. S. & Dixon, N. E. NMR structure of the N-terminal domain of *E. coli* DnaB helicase: implications for structure rearrangements in the helicase hexamer. *Structure* **7**, 681–690 (1999).
43. Slupsky, C. M. *et al.* Structure of the Ets-1 pointed domain and mitogen-activated protein kinase phosphorylation site. *Proc. Natl Acad. Sci. USA* **95**, 12129–12134 (1998).
44. Rhee, S., Martin, R. G., Rosner, J. L. & Davies, D. R. A novel DNA-binding motif in MarA: the first structure for an AraC family transcriptional activator. *Proc. Natl Acad. Sci. USA* **95**, 10413–10418 (1998).

Molecular evidence for genetic mixing of Arctic and Antarctic subpolar populations of planktonic foraminifers

Kate F. Darling^{*†}, Christopher M. Wade^{‡§}, Iain A. Stewart^{*}, Dick Kroon^{*}, Richard Dingle^{||§} & Andrew J. Leigh Brown[†]

^{*} Department of Geology and Geophysics, University of Edinburgh, Edinburgh EH9 3JW, UK

[†] Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, UK

[‡] Institute of Genetics, University of Nottingham, Nottingham NG7 2UH, UK

^{||} British Antarctic Survey, Madingley Road, Cambridge CB3 0ET, UK

Bipolarity, the presence of a species in the high latitudes separated by a gap in distribution across the tropics, is a well-known pattern of global species distribution. But the question of whether bipolar species have evolved independently at the poles since the establishment of the cold-water provinces 16–8 million years ago, or if genes have been transferred across the tropics since that time, has not been addressed. Here we examine genetic variation in the small subunit ribosomal RNA gene of three bipolar planktonic foraminiferal morphospecies. We identify at least one identical genotype in all three morphospecies in both the Arctic and Antarctic subpolar provinces, indicating that trans-tropical gene flow must have occurred. Our genetic analysis also reveals that foraminiferal morphospecies can consist of a complex of genetic types. Such occurrences of genetically distinct populations within one morphospecies may affect the use of planktonic foraminifers as a palaeoceanographic proxy for climate change and necessitate a reassessment of the species concept for the group.

Patterns of bipolar (anti-tropical) distributions are found in a diverse range of terrestrial and marine groups^{1,2}. Bipolar organisms were first observed by biologists during the mid-nineteenth century^{3–5}, and are now known to be recurrent phenomena through considerable spans of geological time¹. Within the marine environment, bipolar species are observed in many planktonic groups. It remains unclear whether they are isolated within their respective high-latitude water masses, or whether trans-tropical transit occurs; such transit would allow intermixing and consequent gene flow between the populations from the northern and the southern hemispheres. Morphology alone can provide few clues, as morphological identity may be due to convergent or parallel evolution in similar environments. Molecular data enable us to investigate this issue directly. If high-latitude populations of bipolar taxa were isolated from one another, different mutations would be expected to accumulate over time, leading to genetic divergence between the two populations. Conversely, if trans-tropical mixing occurs resulting in genetic exchange, then the two populations would be expected to be genetically homogeneous.

Advances in planktonic foraminiferal molecular genetic analysis^{6,7} provide a tool with which to investigate genetic interchange in marine planktonic organisms exhibiting a bipolar distribution⁸. The planktonic foraminifera have an outstanding fossil record, and their calcitic shells (tests) form one of the most widely used microfossil assemblages for the reconstruction of past oceanic environments. They are globally distributed, and their component taxa are found within distinct faunal provinces that broadly correspond to the main hydrographic features of the ocean⁸ (Fig. 1). Planktonic foraminiferal nucleotide sequence data now provide a new dimension for the investigation of oceanic gene flow⁹ which will enable us to answer the key question of whether genetic exchange occurs between high-latitude populations.

Molecular evolution of planktonic foraminifera

Small subunit ribosomal DNA (SSU rDNA) sequences are highly variable in planktonic foraminifera, and phylogenetic analysis has revealed that many morphologically defined species (morphospecies) of planktonic foraminifera in fact represent complexes of different and often highly divergent genetic types (genotypes)⁹. Some of these genotypes are now considered to be cryptic sibling species^{6,10,11}, a commonly observed phenomenon amongst marine taxa¹². We have analysed SSU rDNA sequences of cool-water representatives of three planktonic foraminiferal morphospecies—*Globigerina bulloides* d'Orbigny, *Turborotalita quinqueloba* (Natland) and *Neogloboquadrina pachyderma* (Ehrenberg)—that exhibit a predominantly disjunct, bipolar distribution⁸. Arctic specimens were collected from the south of Iceland to the southeast Greenland margin along two transects, and Antarctic specimens were collected from eight sites along a transect between the Falkland Islands and the Antarctic Peninsula (Fig. 1; Methods).

A foraminiferal SSU rDNA phylogeny incorporating these species is shown in Fig. 2. The phylogeny is based on a restricted data set comprising only those nucleotide positions that could be aligned across all 22 genera (see Fig. 2 legend for details). The placement of *G. bulloides* and *T. quinqueloba* within the spinose cluster in the molecular phylogeny is consistent with the evolutionary relationships inferred from palaeontological data^{13,14}. The location of right- and left-coiling *N. pachyderma* within the benthic region of the tree is, however, inconsistent with palaeontological data, providing further evidence that the extant planktonic foraminifera are not monophyletic in origin^{6,7}.

Each of the *G. bulloides*, *T. quinqueloba* and *N. pachyderma* morphospecies consist of a complex of distinct genotypes (Fig. 2 and Table 1; individual genotypes have identical SSU rDNA sequences). Both left- and right-coiling *N. pachyderma* genotypes cluster with *Neogloboquadrina dutertrei* which forms the warm-water member of the morphological complex¹⁵ (Fig. 2). However, the relationships observed within this cluster are surprising. Unexpectedly, *N. pachyderma* left-coiling genotypes cluster together with

[§] Present addresses: Department of Zoology, The Natural History Museum, London SW7 5BD, UK (C.M.W.); Geological Institute, University of Copenhagen, Oster Voldgade 10, DK-1350 Copenhagen K, Denmark (R.D.).

N. dutertrei, which is right-coiling, and away from the *N. pachyderma* right-coiling genotype. Thus the genetic data we report here confirm that coiling direction in *N. pachyderma* is associated with genetic divergence¹⁶ and not with temperature during development.

Bipolar SSU rDNA genotypes in foraminiferal morphospecies

When considering identity between genotypes, we utilize not only the conserved regions used for the phylogenetic analysis but also the remaining ~500 nucleotide sites from the highly informative variable regions (Fig. 3). Such comparisons of the entire SSU rDNA fragment revealed that all three morphospecies examined had at least one identical SSU rRNA genotype in both the Arctic and Antarctic subpolar provinces (Figs 2 and 3, and Table 1). There were two separate bipolar genotypes (types IIa and IIb) in *G. bulloides*, one bipolar genotype (type IIa) in *T. quinqueloba* and one bipolar genotype (type R) in *N. pachyderma* right-coiling.

The *G. bulloides* genotypes, types IIa and IIb, differed from one another at 36 of 980 nucleotide sites (variable and conserved regions; Fig. 3 and data not shown). There were also two other closely related genotypes in the *G. bulloides* cluster. Type IIc was found only in the subantarctic province, and differed from the type IIa bipolar genotype at 31 sites and from the type IIb bipolar genotype at 49 sites (Fig. 3 and data not shown). Type IId was found in the transitional zone⁹, and differed from the type IIa bipolar

genotype at 43 sites and from the type IIb bipolar genotype at 18 sites (Fig. 3 and data not shown). Two further genotypes were also found in *T. quinqueloba* that were closely related to the bipolar type IIa: type IIb in the subarctic and type IIc in the subantarctic provinces. Type IIb differs from type IIc in only 8 of 1,168 nucleotide sites (variable and conserved) and these differ from the bipolar Type IIa genotype at 83 and 82 nucleotide sites, respectively. The Antarctic type III and type IV genotypes of *N. pachyderma* (left-coiling) differ from one another by as many as 145 nucleotide sites (Fig. 3 and data not shown). Thus, where detailed comparisons may be made across both conserved and variable regions, there is a continuum of genetic divergence in the cool-water 'genotype clusters' extending from identical bipolar genotypes through very closely related variants to others which differ at more than 10% of sites in the region sequenced.

Genetic exchange between high-latitude populations

Foraminiferal sequence data have allowed us to address the fundamental question of whether gene flow occurs between bipolar-distributed planktonic marine organisms. The presence of identical foraminiferal sequences in the high latitudes of each hemisphere, coupled with their extensive fossil record, has allowed us to establish that trans-tropical genetic exchange has been recent. The well-calibrated foraminiferal biostratigraphic record permits the estimation of the expected level of divergence between two sequences over time in the absence of gene flow. Rates of evolution for foraminiferal SSU rRNA genes have been estimated by both our group⁹ and others^{17,18}. These estimates have given approximate rates of substitution of $(2.6\text{--}4.6) \times 10^{-9}$ substitutions per site per year for the spinose planktonic group^{9,18}. Taking an average, we would expect 1.82 substitutions per million years in the conserved 505-base-pair (505-b.p.) region alone, or 29–15 substitutions since the cold-water provinces were established 16–8 million years ago in the Middle to Late Miocene epoch¹⁹.

In addition to the conserved regions, intra-specific comparisons can be made on the variable regions that cannot be utilized for between-morphospecies analysis (Fig. 2). On the basis of our data, the rate of substitution over the variable regions is estimated to be at least 10 times greater than the rate for the conserved regions: thus we expect between 160 and 320 substitutions over the entire sequence since the establishment of the polar provinces. These calculations establish that a substantial degree of divergence would be expected if sequences were isolated for between 8 and 16 million years, yet we have identified examples of complete sequence identity between the high-latitude morphospecies. This strongly suggests that genetic exchange has occurred recently relative to the establishment of the cold-water provinces. This conclusion is reinforced by the fact that it was observed independently in three separate lineages.

Trans-tropical gene flow

Genetic exchange between the high latitudes can only be explained by trans-equatorial transit. The tropical ocean is, however, a formidable barrier for cool-water genotypes to cross (Fig. 1). Foraminiferans do not encyst—a process that provides resistance to inhospitable environments in many other planktonic groups. It is therefore unlikely that the subpolar genotypes would tolerate the high surface water temperatures of the western tropical regions. Yet for complete bipolar genetic mixing to occur, foraminifers crossing the tropics must also return to the high latitudes within the warm tropical ocean gyral surface currents (Fig. 1).

How then can gene flow occur between the cooler-water populations across the tropical Atlantic? The present-day oceanographic setting provides a perspective in which to investigate this issue. The eastern boundaries of the subtropical Atlantic Ocean off west Africa are associated with cool boundary currents (shown as B and C in Fig. 1) which act as corridors for the introduction of cool-water

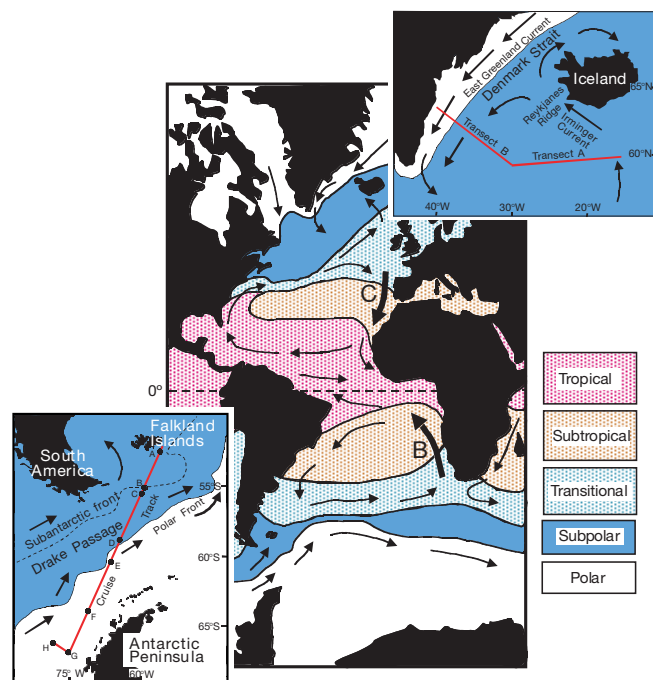


Figure 1 Sampling localities and distribution of the high-latitude morphospecies. The pictorial map (main figure) shows the five zoogeographical planktonic foraminiferal faunal provinces⁸—polar, subpolar, transitional, subtropical and tropical—of the North and South Atlantic Ocean. The three planktonic foraminiferal morphospecies *G. bulloides*, *T. quinqueloba* and *N. pachyderma* (right- and left-coiling) used in this study are found predominantly in the polar, subpolar and transitional provinces, and consequently exhibit a bipolar (anti-tropical) distribution⁹. The cool-water eastern boundary cold currents, the Canary and Benguela (C and B, large arrows), strengthen seasonally and flow equatorwards, bringing cool water and transitional morphospecies into the subtropical province where they bloom in the transient cool seasonal upwelling systems of the subtropical/tropical eastern Atlantic. However, there is always a clear belt of warm tropical water moving seasonally north and south as the seasons build and decay. An outline of the present-day Atlantic Ocean surface circulation patterns (small arrows) are shown. The expanded maps (insets) highlight the sampling transect, sample sites (see Methods) and surface currents within the Arctic and Antarctic subpolar/polar provinces.

genotypes into the cool seasonal upwelling zones of the subtropical region, where they may bloom in significant numbers. The upwelling cells could then provide a 'stepping stone' for the transit of the cool-water genotypes into the permanent equatorial upwelling zone (which is 2–9 °C cooler than surrounding surface water). Yet genetic exchange between populations within these waters alone cannot

account for the genetic homogeneity observed between the subpolar populations. For bipolar genetic exchange to occur, the foraminifera would have to negotiate the high tropical sea surface temperatures of the western equatorial region into which they are passively carried. This region is most unlikely to be conducive to the survival of the cool-water genotypes, either as a direct result of temperature

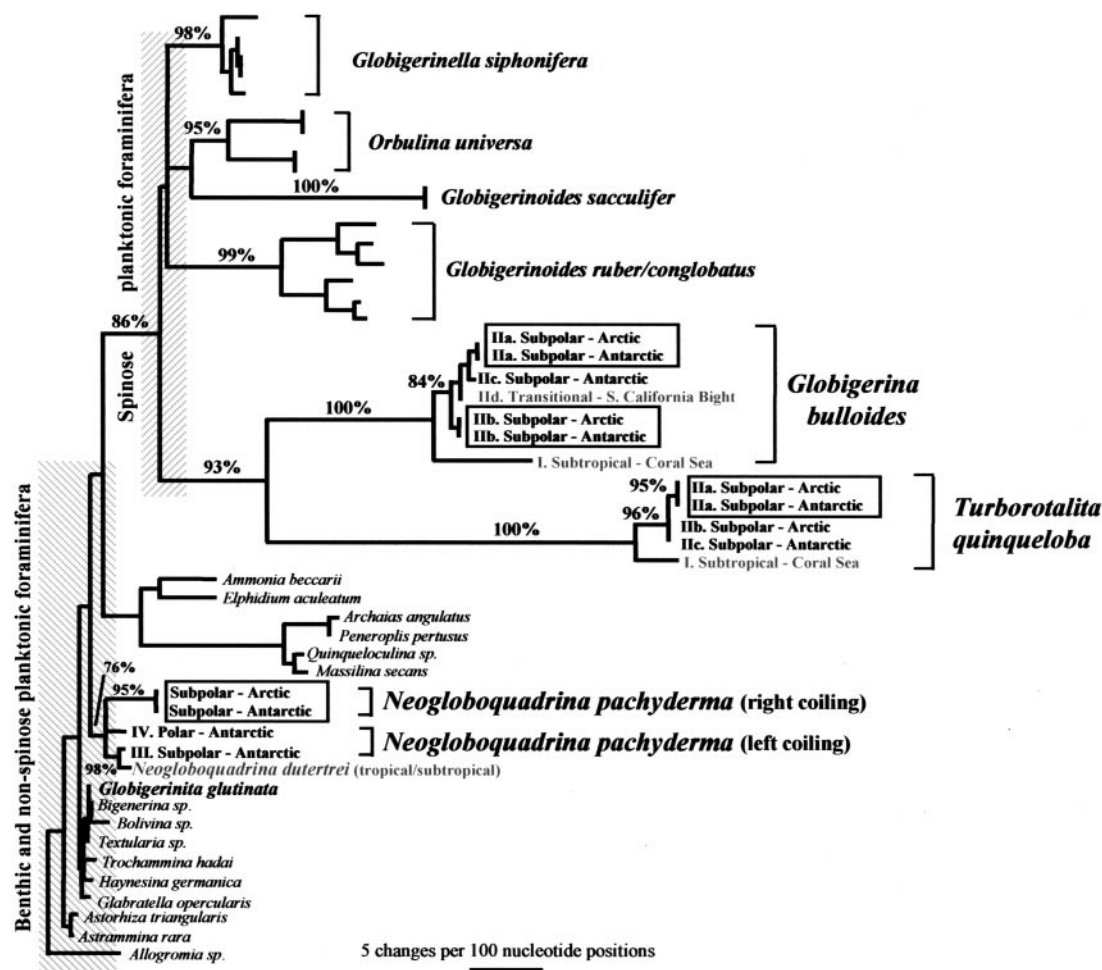


Figure 2 Foraminiferal SSU rDNA phylogeny, highlighting the evolutionary placement of the planktonic foraminiferal high-latitude morphospecies. The phylogeny shows the genotypic variants identified for *G. bulloides* d'Orbigny ($n = 55$), *T. quinqueloba* (Natland) ($n = 24$) and *N. pachyderma* (Ehrenberg) right-coiling ($n = 26$) and left-coiling ($n = 31$). *N. pachyderma* falls within the benthic and non-spinose planktonic foraminiferal region of the tree, and *G. bulloides* and *T. quinqueloba* fall within the spinose planktonic foraminiferal group (shaded boxes). Genotypes of *G. bulloides* and *T. quinqueloba* isolated from the transitional and subtropical regions are shown in grey. *N. dutertrei*, the warm-

water member of the *Neogloboquadrina* morphological complex is also shown in grey. High-latitude genotypes of *G. bulloides*, *T. quinqueloba* and *N. pachyderma* (right- and left-coiling) are shown in black. Genotypes which are identical in both the Arctic and Antarctic (bipolar genotypes) are boxed. Bootstrap values (expressed as a percentage) are shown for the main branches when supported in >70% of replicates. Interrelationships within the transitional and subpolar *G. bulloides* cluster are unresolved. The number of specimens obtained for each genotype and their location (Fig. 1) are shown in Table 1.

Table 1 Number and location of specimens obtained for each genotype.

Morphospecies	Location	Genotype	Specimens	Sample Sites
<i>G. bulloides</i>	Subpolar-Arctic ($n = 44$)	IIa	32	Subarctic transect B
		IIb	12	Subarctic transect A
	Subpolar-Antarctic ($n = 11$)	IIa	5	Site A + B + C + D
		IIb	4	Site D
		IIc	2	Site A
<i>T. quinqueloba</i>	Subpolar-Arctic ($n = 17$)	IIa	8	Subarctic transect A + B
		IIb	9	Subarctic transect A + B
	Subpolar-Antarctic ($n = 7$)	IIa	5	Site C + D
		IIc	2	Site A
<i>N. pachyderma</i>	Subpolar-Arctic ($n = 24$)	Right	24	Subarctic transect A + B
		Right	2	Site A
	Subpolar-Antarctic ($n = 14$)	Left III	12	Site B + C + D
		Left III	3	Site E + F
	Polar-Antarctic ($n = 19$)	Left IV	16	Site E + F + G + H

Locations identified in Fig. 1.

or as a consequence of other physical/biological requirements being limiting within the ecosystem. It is possible that transit occurs through these regions by tropical submergence into cooler levels of the thermocline. The route and timing of transit is likely to be morphospecies-specific due to differences in their capacity to survive transit across the tropics. Whether gene flow is uni-directional or bi-directional remains to be determined.

Trans-equatorial transit and subsequent genetic exchange is likely to increase substantially during cooling periods associated with glacial cycling in the Quaternary period (the past 1.8 million years), and indeed the low-latitude sedimentary record shows the frequent occurrence of planktonic foraminiferal subpolar assemblages within the equatorial zone during these cooling cycles²⁰. Revised estimates of tropical cooling during the Last Glacial Maximum (18,000 ¹⁴C years before present)^{21–24} indicate a substantial drop in temperature of 3–5 °C in the tropical Atlantic, accompanied by an equatorward extension and intensification of the cool-water boundary currents^{25–27} (B and C in Fig. 1). This would have increased the potential for genetic exchange between the bipolar populations. It is, however, possible that genetic exchange could be continuous and may be occurring at present. The available data are unable to resolve timescales of this magnitude, and the issue can only be addressed by sampling living planktonic foraminifera in the subtropical/tropical regions to establish the distribution pattern, proportion and seasonality of cool-water genotypes in the present day.

Speciation in planktonic foraminifers

Whereas molecular studies have shown that many planktonic foraminiferal morphospecies comprise more than one genetically distinct population which may warrant 'cryptic' species status^{6,9,10,11}, we have also shown that planktonic foraminiferal populations intermix on a global scale (Arctic–Antarctic, this study; Pacific–Atlantic, refs 6 and 9). This makes any simple speciation model based on geographical isolation (allopatric speciation) difficult to sustain. Yet allopatric isolation may occur in the marine environment, and there are many examples of population subdivision in marine species despite a high dispersal potential²⁸. Subdivision may result from the sporadic and discontinuous seeding of seasonal upwelling cells or alternating ocean circulation patterns associated with past climate change. Such isolation could give rise to genetically distinct populations.

In order for speciation to occur, genetic divergence must be accompanied by reproductive isolation²⁹. In marine organisms with broadcast release of gametes such as the planktonic foraminifera,

this may be biological, involving, for example, synchrony of gametogenesis^{10,28,30}. Alternatively, isolation may be molecular and there is evidence that this may be achieved through high rates of evolution in proteins involved in gamete recognition^{31,32}. Recent developments in population genetic theory of sympatric speciation^{33,34} suggest further possibilities for speciation in the marine planktonic environment without the necessary initial requirement for allopatry. We conclude that there are several currently recognized mechanisms that could be involved in the speciation of planktonic foraminifera, but the main problem with interpretation of these processes is lack of knowledge of the structuring of their environment in either biotic or abiotic terms. The discovery of genotype complexes in the planktonic foraminifera does, however, necessitate the urgent reassessment of species concepts for the group.

Distribution of planktonic foraminiferal cool-water genotypes

The high-latitude genotypes are not ubiquitous in the surface waters throughout the cool-water provinces. For example, in the Arctic subpolar region, *G. bulloides* type IIa was largely confined to transect B and type IIb to transect A (Fig. 1, Table 1). Given that sampling density was relatively low in the Antarctic subpolar region (due to adverse weather conditions during sampling), *G. bulloides* type IIa was distributed from sites A–D—from the Falkland Islands to the polar front—but type IIb was found only at the polar front (site D, Fig. 1, Table 1). Similarly, *N. pachyderma* left-coiling type III was mostly found between the Falkland Islands and the polar front (sites B–D, Table 1) and type IV was isolated in the true polar waters (sites E–H, Table 1). Although we have no direct evidence at present, such distribution patterns may indicate that individual genotypes are adapted to specific hydrographic or trophic environments; we note that de Vargas *et al.*¹¹ found a close correlation between the distribution of sibling species and trophic regimes in the spinose species *Orbulina universa*.

Implications for palaeoceanography and palaeoclimate

The morphological, chemical and stable-isotope differences associated with planktonic foraminiferal calcitic shells are used extensively by palaeoceanographers for climate reconstruction. For such studies, the assumption is made that each morphospecies represents a genetically continuous species with a single environmental/habitat preference. If this is not the case—as indicated by this study and others^{6,9,10,11}—stable-isotope and geochemical analyses of planktonic foraminiferal shells, and census-based transfer-function techniques derived from such pooled data, must include significant

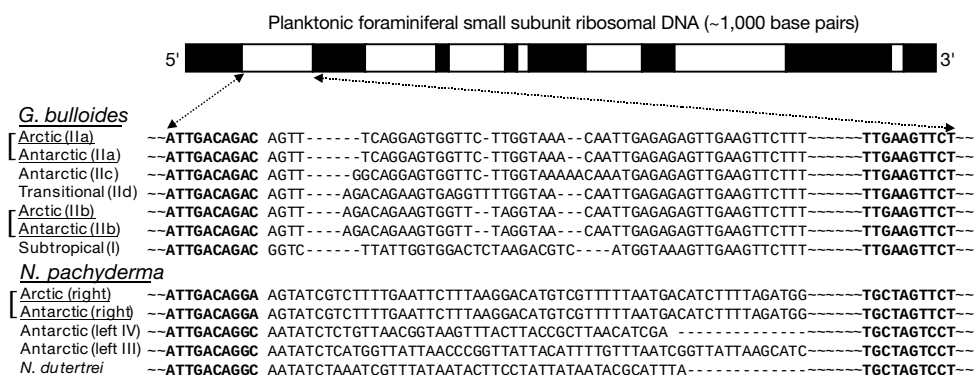


Figure 3 Sequence variability in the SSU rRNA gene. Shown is a schematic representation of the ~1,000-b.p. region of the SSU rRNA gene sequenced for the planktonic foraminifera. Conserved regions alignable across all foraminiferal taxa (505 b.p.) and used in phylogenetic tree construction (see Fig. 2) are highlighted in black. Variable regions, not used in phylogenetic tree construction because they could not be aligned across all taxa, are highlighted in white. Examples of the sequence variation observed in one of the

variable regions are given for *G. bulloides* and *N. pachyderma/dutertrei*. Considerable variability is observed within the variable region, even among closely related cool-water genotypes. The bipolar genotypes (marked with a bracket and underlined) are identical across the entire sequenced region. (~ denotes sequences not displayed; – denotes gaps introduced in aligning sequences).

noise, if not error. If genetic differences can be correlated with specific environment and habitat preferences, and the genotypes differentiated in the fossil record, a new level of precision for climate modelling could be achieved. □

Methods

Sampling localities

Arctic subpolar specimens of *G. bulloides*, *T. quinqueloba* and *N. pachyderma* right-coiling were collected on board RV *Professor Logachev* along two transects (A and B, Fig. 1). Transect A ran from a point 59° 58' N, 11° 34' W (south of Iceland) to 58° 56' N, 30° 24' W (above the Reykjanes ridge). Transect B ran from the Reykjanes ridge to the southeast Greenland margin at 63° 48' N, 40° 20' W across the Denmark Strait. Specimens were obtained by pumping continually from 4.5 m depth through a plankton net suspended on deck. Antarctic subpolar specimens of *G. bulloides*, *T. quinqueloba* and *N. pachyderma* left- and right-coiling were collected from eight sites (A–H) between the Falkland Islands (53° 21' S, 58° 20' W) and west of the Antarctic Peninsula (65° 36' S, 77° 39' W) (inset map, Fig. 1). Specimens were obtained by pumping from 6 m depth through a 63-µm filter on board RRS *James Clark Ross* (JR19). Transitional specimens of *G. bulloides* were collected ~2 km NNE of the Santa Catalina Marine Science Centre, Santa Catalina Island, California (33° N, 118° W), and subtropical specimens of *G. bulloides* and *T. quinqueloba* were collected off the Great Barrier Reef, Australia (0.8 nautical miles due east of Ribbon Reef 10).

Isolation and sequencing of SSU genes

DNA extraction, amplification by polymerase chain reaction and direct automated sequencing of an ~1,000-b.p. region of the terminal 3' end of the foraminiferal SSU rRNA gene was as described previously^{6,9,35}. The SSU rRNA genes form a large multigene family in all organisms where data are available. Within the gene family, as mutations accumulate in individual copies, a level of homogeneity is maintained between copies by the process of concerted evolution. All data presented here have been derived from a consensus sequence amplified from the gene family of a single individual using a direct sequencing approach. This has the advantage of eliminating the risk of sampling non-orthologous copies, as different copies would be detected as ambiguities in the consensus sequence. Using this approach we have detected no evidence of ambiguity in any of the genotypes presented here, apart from very minor variability between copies within individuals of left-coiling *N. pachyderma*.

Sequence analysis

Partial SSU rDNA sequences were aligned manually within version 2.2 of the Genetic Data Environment (GDE) package³⁶. 505 unambiguously aligned nucleotide sites were employed in phylogeny reconstruction. The regions included are indicated schematically in Fig. 3 ("conserved regions"). A neighbour-joining³⁷ phylogeny was generated using γ rate corrected F84 distances within PAUP* version 4.0d64 (kindly provided by D. L. Swofford³⁸). α values for rate correction were estimated within PAUP*. Bootstrap resampling³⁹ (2,000 replicates) was employed to assign support to the neighbour-joining tree. Cool-water planktonic foraminiferal taxa are placed within a background of additional planktonic⁴⁰ and benthic¹⁷ foraminifers. The tree was rooted on *Allogromia*, a membranous-walled benthic foraminifer which is thought to have diverged early in the history of the clade⁴⁰.

Received 19 April 1999; accepted 24 February 2000.

- Crame, J. A. Bipolar molluscs and their evolutionary implications. *J. Biogeogr.* **20**, 145–161 (1993).
- Lindberg, D. R. Marine biotic interchange between the northern and southern hemispheres. *Paleobiology* **17**, 308–324 (1991).
- Darwin, C. *The Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (John Murray, London, 1859).
- Hooker, J. D. *The Botany: The Antarctic Voyage of H. M. Discovery Ships Erebus and Terror, in the Years 1839–1843* (Reeve Brothers, London, 1847).
- Ross, J. C. *A Voyage of Discovery and Research in the Southern and Antarctic Regions, during the Years 1839–43* (John Murray, London, 1847).
- Darling, K. F., Wade, C. M., Kroon, D. & Leigh Brown, A. J. Planktic foraminiferal molecular evolution and their polyphyletic origins from benthic taxa. *Mar. Micropaleontol.* **30**, 251–266 (1997).
- de Vargas, C., Zaninetti, L., Hilbrecht, H. & Pawlowski, J. Phylogeny and rates of molecular evolution of planktonic foraminifera: SSU rDNA sequences compared to the fossil record. *J. Mol. Evol.* **45**, 285–294 (1997).
- Bé, A. W. H. in *Oceanic Micropaleontology* (ed. Ramsey, A. T. S.) 1–100 (Academic, London, 1977).
- Darling, K. F., Wade, C. M., Kroon, D., Leigh Brown, A. J. & Bijma, J. The diversity and distribution of modern planktic foraminiferal small subunit ribosomal RNA genotypes and their potential as tracers of present and past ocean circulations. *Paleoceanography* **14**, 3–12 (1999).
- Huber, B. T., Bijma, J. & Darling, K. F. Cryptic speciation in the living planktonic foraminifer *Globigerinella siphonifera* (d'Orbigny). *Paleobiology* **23**, 33–62 (1997).

- de Vargas, C., Norris, R., Zaninetti, L., Gibb, S. W. & Pawlowski, J. Molecular evidence of cryptic speciation in planktonic foraminifers and their relation to oceanic provinces. *Proc. Natl Acad. Sci. USA* **96**, 2864–2868 (1999).
- Knowlton, N. Sibling species in the sea. *Annu. Rev. Ecol. Systematics* **24**, 189–216 (1993).
- Kennett, J. P. & Srinivasan, M. S. *Neogene Planktonic Foraminifera, a Phylogenetic Atlas* (Hutchinson and Ross, Stroudsburg, Pennsylvania, 1983).
- Spezzaferri, S. Planktonic foraminiferal biostratigraphy of the Oligocene and lower Miocene in the oceanic record. An overview. *Palaeontogr. Italica* **81**, 1–187 (1994).
- Hilbrecht, H. Morphologic gradation and ecology in *Neogloboquadrina pachyderma* and *N. dutertrei* (planktic foraminifera) from core top sediments. *Mar. Micropaleontol.* **31**, 31–43 (1997).
- Brummer, G. A. & Kroon, D. in *Planktonic Foraminifers as Tracers of Ocean-climate History* (eds Brummer, G. A. & Kroon, D.) 293–298 (Free University, Amsterdam, 1988).
- Pawlowski, J. et al. Extreme differences in rates of molecular evolution of foraminifera revealed by comparison of ribosomal DNA sequences and the fossil record. *Mol. Biol. Evol.* **14**, 498–505 (1997).
- de Vargas, C. & Pawlowski, J. Molecular versus taxonomic rates of evolution in planktonic foraminifera. *Mol. Phylog. Evol.* **9**, 463–469 (1998).
- Kennett, J. P., Keller, G. & Srinivasan, M. S. Miocene planktonic foraminiferal biogeography and paleoceanographic development of the Indo-Pacific region. *Geol. Soc. Am.* **163**, 197–236 (1985).
- McIntyre, A., Ruddiman, W. F., Karlin, K. & Mix, A. C. Surface water response of the equatorial Atlantic Ocean to orbital forcing. *Paleoceanography* **4**, 19–55 (1989).
- Guilderson, T. P., Fairbanks, R. G. & Rubenstone, J. L. Tropical temperature variations since 20,000 years ago: modulating interhemispheric climate change. *Science* **263**, 663–665 (1994).
- Stute, M. et al. Cooling of tropical Brazil (5°C) during the last glacial maximum. *Science* **269**, 379–383 (1995).
- Schrag, D. P., Hampt, G. & Murray, D. Pore fluid constraints on the temperature and oxygen isotopic composition of the glacial ocean. *Science* **272**, 1930–1932 (1996).
- Wolff, T., Mülitz, S., Arz, H., Pätzold, F. & Wefer, G. Oxygen isotopes versus CLIMAP (18ka) temperatures: a comparison from the tropical Atlantic. *Geology* **26**, 675–678 (1998).
- CLIMAP Project Members. Seasonal reconstructions of the Earth's surface at the last glacial maximum. *Geol. Soc. Am. Map Chart Ser.* **MC-36**, 1–18 (1981).
- Mix, A. C. & Morey, A. E. in *The South Atlantic: Present and Past Circulation* (eds Wefer, G., Berger, W. H., Siedler, G. & Webb, D. J.) 503–525 (Springer, Berlin, 1996).
- Zhao, M., Beveridge, N. A. S., Shackleton, N. J., Sarnthein, M. & Eglinton, G. Molecular stratigraphy of cores off northwest Africa: Sea surface temperature history over the last 80 ka. *Paleoceanography* **10**, 661–675 (1995).
- Palumbi, S. R. Genetic divergence, reproductive isolation, and marine speciation. *Annu. Rev. Ecol. Systematics* **25**, 547–572 (1994).
- Coyne, J. A. & Orr, H. A. The evolutionary genetics of speciation. *Phil. Trans. R. Soc. Lond. B* **353**, 287–305 (1998).
- Bijma, J., Erez, J. & Hemleben, C. Lunar and semi-lunar reproductive cycles in some spinose planktonic foraminifers. *J. Foramin. Res.* **20**, 117–127 (1990).
- Metz, E. C. & Palumbi, S. R. Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. *Mol. Biol. Evol.* **13**, 397–406 (1996).
- Biermann, C. H. The molecular evolution of sperm bindin in six species of sea urchins (Echinoidea: Strongylocentrotidae). *Mol. Biol. Evol.* **15**, 1761–1771 (1998).
- Dieckmann, U. & Doebeli, M. On the origin of species by sympatric speciation. *Nature* **400**, 354–357 (1999).
- Tregenza, T. & Butlin, R. K. Speciation without isolation. *Nature* **400**, 311–312 (1999).
- Darling, K. F., Kroon, D., Wade, C. M. & Leigh Brown, A. J. Molecular evolution of planktic foraminifera. *J. Foramin. Res.* **26**, 324–330 (1996).
- Smith, S. W., Overbeek, R., Woese, C. R., Gilbert, W. & Gillevet, P. M. The genetic data environment, an expandable GUI for multiple sequence analysis. *Comput. Appl. Biosci.* **10**, 671–675 (1994).
- Saitou, N. & Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
- Swofford, D. L. PAUP*. Phylogenetic Analysis using Parsimony (*and Other Methods) V. 4 (Sinauer Associates, Sunderland, Massachusetts, 1998).
- Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791 (1985).
- Tappan, H. & Loeblich, A. R. Foraminiferal evolution, diversification and extinction. *J. Paleontol.* **62**, 695–714 (1988).

Acknowledgements

Acknowledgements We thank C. Pudsey and T. Paramour for their assistance during the 1997 Antarctic cruise on RRS *James Clark Ross* (British Antarctic Survey), and S. Troelstra and colleagues for their support during the 1997 Denmark Strait expedition on RV *Professor Logachev* (Netherlands Foundation for Scientific Research and the Geological Survey of Denmark and Greenland). We also thank A. Davison, S. Goodacre, J. Brookfield, P. Sharp and B. Clarke for discussions and comments on the manuscript. This work was supported by the Leverhulme Trust, the Carnegie Trust for the Universities of Scotland and the NERC.

Correspondence and requests for materials should be addressed to K.F.D. (e-mail: Kate.Darling@ed.ac.uk). The sequences determined for this study are available in GenBank under the accession numbers AF250105–AF250120.

Io as a source of the jovian dust streams

A. L. Graps*, E. Grün*, H. Svedhem†, H. Krüger*, M. Horányi‡, A. Heck* & S. Lammers*

* Max-Planck-Institut für Kernphysik, Saupfercheckweg 1, 69117 Heidelberg, Germany

† European Space Research and Technology Centre, 2200 AG Noordwijk, The Netherlands

‡ Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, Colorado 80309, USA

Streams of dust emerging from the direction of Jupiter were discovered in 1992 during the flyby of the Ulysses spacecraft^{1,2}, but their precise origin within the jovian system remained unclear². Further data^{3–5} collected by the Galileo spacecraft, which has been orbiting Jupiter since December 1995, identified the possible sources of dust as Jupiter's main ring⁶, its gossamer ring⁷, comet Shoemaker–Levy 9 (ref. 8) and Io. All but Jupiter's gossamer ring and Io have since been ruled out^{4,9–14}. Here we find that the dominant source of the jovian dust streams is Io, on the basis of periodicities in the dust impact signal. Io's volcanoes, rather than impact ejecta, are the dust sources.

The jovian dust stream population detected in the inner satellite region by the Galileo dust detector system (DDS) can be seen in the data's highly variable (over periods of hours) impact rates of submicrometre-sized particles. Using a Lomb–Scargle periodogram¹⁵, we have transformed, from the time domain to the frequency domain, the first two years (1996–1997) of Galileo DDS impact-rate data while in orbit around Jupiter. Eight peaks are prominent in the frequency range of 0 to 6 cycles d^{-1} (Fig. 1). We interpret these frequency peaks on the basis of the following description of our physical system (see Fig. 2).

Io, with its ~ 42 -h rotation period, provides material from its volcanic plumes (up to ~ 460 km in height¹⁶). Material escapes Io at an approximate rate of 1 ton s^{-1} (ref. 17), through a multistep process involving Io's atmosphere and the local plasma environment¹⁸. If the charge is high enough, some material can be swept up by Jupiter's magnetic field, which orbits with Jupiter with a ~ 10 -h rotation period^{19,20}.

Galileo's and Io's orbits lie very nearly in Jupiter's equatorial plane (within 1°). Owing to a 10° tilt between Jupiter's rotational and magnetic axis, Io and Galileo pierce Jupiter's magnetic equatorial plane twice each orbit. Because charged dust couples to Jupiter's magnetic field, which co-rotates with Jupiter (frequency 2.4 d^{-1}), Jupiter modulates the rate at which dust particles emerging from Io (frequency 0.6 d^{-1}) can be detected. Galileo's dust detector records these charged particles when Jupiter's warped dust sheet passes over its position, which occurs on average twice per Jupiter rotation (4.8 d^{-1}).

We interpret the frequency peaks seen in Fig. 1 to be the result of Io's frequency of rotation, Jupiter's magnetic field frequency of rotation and an interaction between these two frequencies called amplitude modulation. The simplest case of amplitude modulation is a sinusoid modulating the amplitude of a carrier signal, which is itself a sinusoid. Then the carrier signal is broken down in frequency space into several sinusoidal oscillations: $x \approx \sin(\omega_0 t) + \sin(\omega_0 t)\sin(\Omega t)$, which can be converted to sums of frequencies using a trigonometric identity for sine products. The result is a signature in frequency space that displays a carrier frequency, ω_0 , with side frequencies ('modulation products'), $(\omega_0 + \Omega)$ and $(\omega_0 - \Omega)$. Jupiter's modulation of Io's frequency signal can be seen as sidelobes around Jupiter's rotation frequency ($\Omega_0 = 2.4 \text{ d}^{-1}$), with the pattern repeated at the first harmonic of

Jupiter's rotation frequency (4.8 d^{-1}).

The presence of Io's rotation frequency suggest that Io is a localized source of charged dust particles because charged dust from diffuse sources would couple to Jupiter's magnetic field and appear in frequency space with Jupiter's rotation frequency and its harmonics. A confirmation of Io's role as a localized charged dust source arises through the modulation effects.

Several Galileo spacecraft orbital characteristics can also be identified in the frequency-transformed data. For each orbit, the dust detector receives more dust impacts in the inner jovian system than in the outer jovian system, which results in a gradual rise and fall in the number of dust impacts. In frequency space, such a 'hump' manifests itself as a low-frequency event close to the origin at Galileo's orbital frequency. Also for each orbit, because Galileo and Io, as the 'observer' and the 'source', are travelling towards and away from each other, Doppler effects shift the signal to higher and lower frequencies; however, we do not see the long periods because of the detector's field of view. Instead, we see Io's orbital frequency Doppler-shifted to shorter periods by roughly 1–2 h, depending on the reference speed of the dust streams, and we believe that this Doppler smearing results in the asymmetry of the Io frequency peak, seen in Fig. 1.

We can immediately perform two simple calculations to investigate Io's role as a source of the dust streams: first, we can find where Io is physically located at the time of a source particle's dynamical trajectory; and second, we can find the amount of dust needed from Io to explain the dust measurements. For our first calculation, using 1996–1998 data of dust impact rates, we found that Io was usually physically located within 50° longitude of the origin of the particle's model trajectory. For our second calculation, if we assumed a wedge-shaped emission pattern of dust stream particles detected

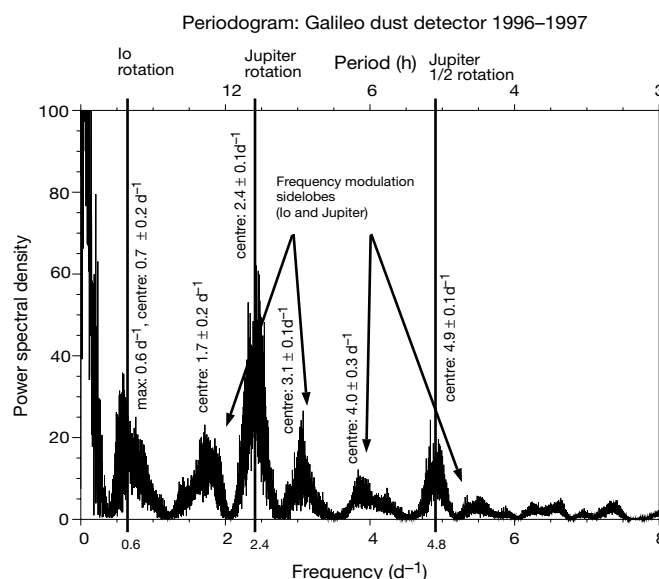


Figure 1 A Lomb–Scargle periodogram¹⁵ for the first two years, 1996–1997, of Galileo data on the dust impact rate. We see the following frequencies in this periodogram: (1) a strong peak near the origin; (2) an asymmetric peak, maximum at 0.6 d^{-1} , centre at $0.7 \pm 0.2 \text{ d}^{-1}$; (3) an asymmetric peak, centre at $1.7 \pm 0.2 \text{ d}^{-1}$; (4) a tall peak, centre at $2.4 \pm 0.1 \text{ d}^{-1}$; (5) a peak, centre at $3.1 \pm 0.1 \text{ d}^{-1}$; (6) harmonics of the previous three peaks; and (7) progressively smaller and less-defined peaks. Vertical solid lines mark Io's and Jupiter's rotational periods (see ref. 23), and arrows point to Jupiter's modulation products with Io straddling Jupiter's frequency. The first harmonic of Jupiter's rotation frequency is visible at ($\omega_1 = 4.8 \text{ d}^{-1}$) and Jupiter's modulation products with Io, which are straddling that first harmonic peak, can also be seen. The strong frequency peak near the origin at 1 over Galileo's orbital period is due to the Galileo spacecraft orbital geometry.

by Galileo, a 2 g cm^{-3} particle density, a 10-nm particle radius, a 0.0235-m^2 detector sensitivity, an $11\text{--}45^\circ$ wedge originating at Jupiter, a $30\text{-}R_J$ (where $R_J = 7.1 \times 10^4 \text{ km}$, the radius of Jupiter) distance between Galileo and Jupiter, and an average DDS rate of $0.2\text{--}30 \text{ particles min}^{-1}$ from the larger 1996–1999 DDS impact-rate dataset, we found that the total mass rate was 10.0 g s^{-1} to 10.0 kg s^{-1} . The periodogram peaks gave us a rough indication of the fraction of dust from Io, if we summed under the Io peak and the modulation sidelobes (primary and first harmonic). That fraction amounted to 60.0% of the total mass rate, or 6.0 g s^{-1} to 6.0 kg s^{-1} . The remaining mass fraction could come from Io or elsewhere, but it has lost Io's frequency signature.

For the gossamer ring to be a direct source of the jovian dust streams, the dynamical requirements have been shown¹² to be that the initial particle size distribution must be within a narrow size range and the charged dust particles must have enough energy to break through the cold plasma torus, which is located slightly inside of the orbit of Io. The size range required was 50–100 nm, because smaller particles than this would be bent away from the equatorial plane and not be able to reach Ulysses and Galileo on approach, and larger particles than these would move inward because of Poynting–Robertson drag. Note that this size range is in disagreement with the 5–10-nm sized particles required for Ulysses to detect the dust streams in interplanetary space.

Now let us examine the possibility of the gossamer ring or some other non-Io material being an indirect source of the jovian dust streams, with Io modulating this non-Io source material. In this case, some of the same above dynamical constraints apply, and, to

succeed, three intermediate conditions would have to be satisfied. (1) The material would first need to be transported from their source region to the plasma torus by some diffusion mechanism, as yet unidentified. (2) The grains in the right size range would have to be stored in the plasma torus, with their influx rate and their assumed loss balanced to maintain the grains' influx and storage in the torus. (3) Io would have to be able to alter the torus plasma environment significantly to change the 'storage conditions' to 'ejection conditions'. This latter condition is not an easy feat. Io's material injection rate of about 1 ton s^{-1} into the torus, although seeming very large, is a very small mass compared with the total of $10^5\text{--}10^6 \text{ ton}$ in the torus itself. The effect of Io on the torus is therefore very tiny and the probability that any dust stored in the torus may be picking up this effect is negligible. We believe that any of these three conditions individually might be possible (however unlikely), but together they form a compelling argument against Io's ability to modulate non-Io source dust material.

We are therefore left with the much simpler explanation that Io itself is the source of the jovian dust streams. For this scenario to succeed, we need simply the intermediate condition that a small fraction (less than 1%) of the volcanic 1 ton s^{-1} material in the form of 10-nm particles must escape, either directly by entrainment in the plumes, or indirectly as condensates from the expanding, rapidly cooling, escaping gas.

In summary, frequency analysis of the Galileo DDS data provides our first direct evidence of an Io dust source. We believe that the Io dust source is predominately Io's volcanoes, rather than impact ejecta from Io, because the dust stream observations show that the particles are of submicrometre size, and the size of the dust particles that can escape from a typical Io volcanic plume has been shown to be $0.01 \text{ }\mu\text{m}$ or less (ref. 20). In addition, a study has shown that Io impact ejecta is not effective enough, by far, to be a dominant source of dust for the dust streams²¹.

We see several consequences for this discovery. Dust from Io's volcanoes is a minor dust source compared with collisions of the main belt asteroids and comet activity, nevertheless, it adds to the variety of dust sources in the solar system. At a velocity of 200 km s^{-1} or more (ref. 22), the jovian dust stream particles can also leave the solar system to slightly populate the local interstellar medium. We can now use dust stream measurements to monitor Io's volcanoes' plume activity. Such measurements are a unique complement to the partial glimpses acquired by Galileo and ground-based image observations, because our temporal coverage is more complete (we provide information for each Galileo orbit) and because the DDS measurements give an estimate of the integrated total amount of volcanic material from Io's (more than 100) volcanoes, for example, material that escapes from Io, and which then disperses through the jovian system painting the other satellite surfaces.

This discovery also lends support for using dust measurements as a probe for charging effects in the jovian magnetosphere. Dust from the dust streams is clearly magnetically controlled. Dust particles carry information about charging processes in regions of the jovian magnetosphere, where information is otherwise sparse or unknown. And finally, in December 2000, we will have a unique opportunity to measure the elemental composition of the Io dust stream particles, confirming the dust stream origin by Cassini's Cosmic Dust Analyzer simultaneously with the Galileo Dust Detector System during the Cassini Jupiter flyby. □

Received 20 September 1999; accepted 13 March 2000.

1. Grün, E. *et al.* Ulysses dust measurements near Jupiter. *Science* **257**, 1550–1552 (1993).
2. Grün, E. *et al.* Discovery of Jovian dust streams and interstellar grains by the Ulysses spacecraft. *Nature* **362**, 428–430 (1993).
3. Grün, E. *et al.* Dust measurements during the initial Galileo Jupiter approach and Io encounter. *Science* **274**, 399–401 (1996).
4. Grün, E. *et al.* Constraints from Galileo observations on the origin of Jovian dust streams. *Nature* **381**, 395–398 (1996).
5. Grün, E. *et al.* Galileo observes electromagnetically coupled dust in the Jovian magnetosphere. *J. Geophys. Res.* **103**, No. E9, 20, 011–20, 022 (1998).

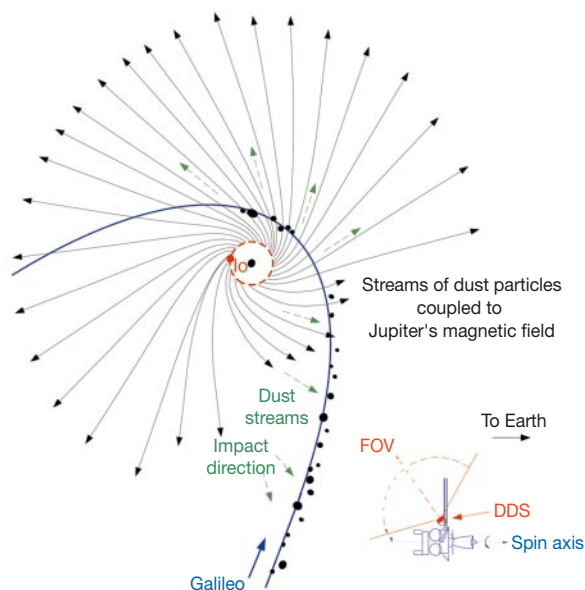


Figure 2 Sketch of one of Galileo's orbital trajectories from late 1997 overlaid with trajectory results from a model of dust stream particles¹². The Galileo orbital trajectory is indicated with a solid blue line and the impact direction of the dust streams is indicated by the dashed green arrows. The Dust Detection System (DDS) instrument is mounted on the spinning portion (rotation angle $0\text{--}360^\circ$) of the spacecraft, its centre line offset from the spacecraft spin axis by 60° and subtending a 140° field of view (FOV). The perijove of the spacecraft's orbital ellipse shifts over time, and, owing to the detector's geometrical orientation, the DDS instrument captures dust stream particles at different locations in its trajectories during the mission. In this sketch from late 1997, the dust stream particles on the approach leg of Galileo's trajectory could enter the dust detector, whereas on the receding leg the dust streams approach Galileo from outside the detector field of view. The DDS instrument (in 1996–1997) therefore captured most of the dust stream particles during the inbound leg of Galileo's orbital trajectory. As Galileo's orbital speed is $\leq 20 \text{ km s}^{-1}$ and the dust streams' speed is 200 km s^{-1} or more (ref. 22), Galileo, to first order, is a stationary observer with respect to the dust streams.

6. Showalter, M. R., Burns, J. A., Cuzzi, J. N. & Pollack, J. B. Jupiter's ring system—New results on structure and particle properties. *Icarus* **69**, 458–498 (1987).
7. Showalter, M. R., Burns, J. A., Cuzzi, J. N. & Pollack, J. B. Discovery of Jupiter's 'gossamer' ring. *Nature* **316**, 526–528 (1985).
8. Sekanina, A., Chodas, P. W. & Yeomans, D. K. Tidal disruption and the appearance of periodic comet Shoemaker–Levy 9. *Astron. Astrophys.* **289**, 607–636 (1994).
9. Grün, E. *et al.* Dust streams from comet Shoemaker–Levy 9? *Geophys. Res. Lett.* **21**, 1035–1038 (1994).
10. Hamilton, D. & Burns, J. A. Ejection of dust from Jupiter's gossamer ring. *Nature* **364**, 695–699 (1993).
11. Horányi, M., Morfill, G. & Grün, E. Mechanism for the acceleration and ejection of dust grains from Jupiter's magnetosphere. *Nature* **363**, 144–146 (1993).
12. Horányi, M., Morfill, G. & Grün, E. The dusty ballerina skirt of Jupiter. *J. Geophys. Res.* **98**, 21,245–21,251 (1993).
13. Horányi, M., Grün, E. & Heck, A. Modeling the Galileo dust measurements at Jupiter. *Geophys. Res. Lett.* **24**, 2175–2178 (1997).
14. Maravilla, D., Flammer, K. R. & Mendis, D. A. On the injection of fine dust from the Jovian magnetosphere. *Astrophys. J.* **438**, 968–974 (1995).
15. Scargle, J. D. Studies in astronomical time series II: Statistical aspects of spectral analysis of unevenly spaced data. *Astrophys. J.* **263**, 835–853 (1982).
16. McEwen, A. S. *et al.* Active volcanism on Io as seen by Galileo SSL. *Icarus* **135**, 181–219 (1998).
17. Spencer, J. R. & Schneider, N. M. Io on the eve of the Galileo mission. *Annu. Rev. Earth Planet. Sci.* **24**, 125–190 (1996).
18. McGrath, M. Io and the plasma torus. *Science* **278**, 237–238 (1997).
19. Johnson, T. V., Morfill, G. & Grün, E. Dust in Jupiter's magnetosphere—an Io source. *Geophys. Res. Lett.* **7**, 305–308 (1980).
20. Ip, W. H. The dust halo of Io. *Geophys. Res. Lett.* **24**, 3671–3674 (1996).
21. Krüger, H., Krivov, A. V., Hamilton, D. P. & Grün, E. Discovery of a dust cloud around Ganymede. *Nature* **399**, 558–560 (1999).
22. Zook, H. A. *et al.* Solar wind magnetic field bending of Jovian dust trajectories. *Science* **274**, 1501–1503 (1996).
23. Carr, T. D., Desch, M. D. & Alexander, J. K. in *Physics of the Jovian Magnetosphere* (ed. Dessler, A. J.) 248 (Cambridge Univ. Press, Cambridge, 1983).

Acknowledgements

We thank the Galileo project at JPL for effective and successful mission operations. We also thank J. A. Burns for helpful comments. This work was supported by Deutsches Zentrum für Luft-und Raumfahrt E.V. (DLR).

Correspondence and requests for materials should be addressed to A.L.G. (e-mail: Amara.Graps@mpi-hd.mpg.de).

Detection of doubled shot noise in short normal-metal/superconductor junctions

X. Jehl^{*}, M. Sanquer^{*}, R. Calemczuk^{*} & D. Mailly[†]

^{*} DRFMC-SPSMS, CEA-Grenoble, F-38054 Grenoble, France

[†] Laboratoire de Microstructures et de Microélectronique, CNRS-LMM, F-92225 Bagneux, France

Shot noise refers to the fluctuations in electrical current through a device arising from the discrete nature of the charge-carrying particles. Recent experiments have exploited the fact that the shot noise is proportional to the charge of the carriers to establish fractional quantization of quasiparticles in the fractional quantum Hall effect^{1,2}. By a similar argument, it is expected that when a superconducting reservoir emits Cooper pairs, (which have a charge twice that of an electron) into a short normal-metal wire, the shot noise should be double that obtained for a normal-metal reservoir^{3,4}. Although the charge of Cooper pairs has been well established by flux quantization and tunnel experiments⁵, doubling of their shot noise has not yet been observed. Here we report a shot-noise experiment using a short diffusive normal-metal superconductor contact, in which we confirm the predicted noise behaviour for double charges. The measurements, taken over a large range of bias current, establish that phase coherence is not required to observe the effect.

In a diffusive metal, where the electronic mean free path l is much smaller than the sample length L , and in the absence of inelastic interactions, shot noise is reduced below the classical Schottky value⁶ for uncorrelated events by a factor of 1/3. This so-called Fano factor $F = S_I/2eI = 1/3$ (where I is the mean bias current, and e is the electron's charge) was simultaneously obtained from coherent scattering⁷ and semiclassical⁸ theories, and does not depend on the geometry of the sample. It has been measured in silver⁹ and gold wires¹⁰. The fact that the same reduction factor 1/3 is obtained by quantum coherent and classical theories shows that the phase coherence of the transmitted electrons is only a second-order effect in the calculation of F (ref. 11).

In normal-metal/superconductor (NS) junctions, the dominant low-energy transport process involves an incident electron from the normal side being reflected as a hole, while an electron pair is absorbed by the superconducting condensate¹². The shot noise in NS contacts is predicted to increase by a factor of two because of the double charge transmitted by these Andreev reflections^{3,4,13}. Previous studies concentrated on superconductor/normal-metal/superconductor junctions, where, if the structure is short enough, multiple Andreev reflections are responsible for a giant excess noise at low bias^{14,15}. In longer samples an increase of the shot noise was observed, but did not reach the predicted factor of two increase¹⁶.

To see unambiguously the doubled Andreev shot noise in a diffusive NS structure, a normal disordered wire of length L must be placed between a superconducting and a normal reservoir. To prevent energy relaxation of electrons towards the local equilibrium along the wire¹⁷, the diffusion time $\tau_D = L^2/D$ (D is the diffusion constant) must be smaller than the electron–electron interaction relaxation time τ_{ee} . Then the Fano factor is predicted to be $F = 2/3$; that is, twice the factor of 1/3 that applies to normal samples⁴. Until now this calculation relied on the coherent scattering formalism, and therefore the effect of interactions—for instance, in altering the distribution function in long NS junctions—was not well established. The measurement requires a clean NS contact, because the number of Andreev-reflected electrons is proportional to the square of the barrier transmission and a barrier at the interface would give

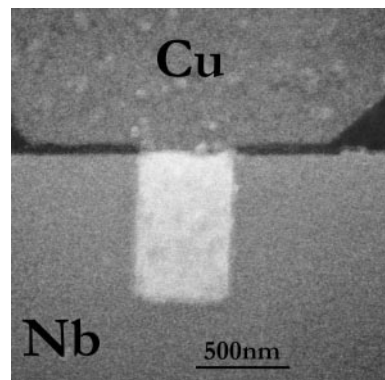


Figure 1 Scanning electron microscope image of the sample. It shows the superconducting niobium (Nb) and normal copper (Cu) reservoirs, with a Cu finger overlapping the Nb defining the junction. The device was fabricated by depositing first a 50-nm-thick Nb film on a Si(100) substrate by argon sputtering. All the subsequent lithography was done using a JEOL 5DIIU electron beam writer on poly(methyl methacrylate) and gold registration marks for the layer alignment. The Nb electrode is defined by SF₆ reactive-ion etching through an aluminium mask, while the Si substrate was also etched to a thickness of 50 nm. After the removal of the aluminium mask with a NaOH solution and another lithography step, the 100-nm-thick Cu electrode is then Joule-evaporated, and lift-off leaves 800 nm × 500 nm Cu finger on the top of the Nb. The good transparency of the junction is obtained by a 500-V argon etching of the Nb before the Cu deposition. The Cu film has a sheet resistance $R_{\square} = 0.15 \Omega$ at $T = 1.4$ K, resulting in a diffusion constant $D \approx 2 \times 10^{-2} \text{ m}^2 \text{ s}^{-1}$ or an elastic scattering length $l \approx 40$ nm. The Nb film is superconducting below $T = 8.3$ K.

too small a current. With a clean NS contact and a short diffusive normal wire, the NS sample has a low resistance. A large current is necessary to distinguish the shot noise ($S_1 \propto I$) from the Johnson–Nyquist thermal current noise ($4k_B T/R$ for a resistance R at temperature T ; k_B is Boltzmann’s constant). A SQUID-based resistance bridge¹⁸ was used to measure the current fluctuations through the device; this gives excellent sensitivity and the ability to sustain large bias currents.

Figure 1 shows a picture of the sample. The total resistance of our NS device was typically 0.8Ω (Fig. 2); this figure includes the resistance of the Nb/Cu interface, the intrinsic resistance of the Cu finger overlapping the Nb, and the resistance of the Cu film at the edge of the Nb electrode. The values of these individual contributions cannot be precisely estimated, but the analysis of the temperature dependence of the resistance below $T = 4.2$ K (see Fig. 2) gives the necessary information on the sample parameters relevant to our experiment whatever the specific geometry of the sample. In the diffusive regime, the Fano factor of $2/3$ does not depend on the geometry.

With decreasing temperature, the resistance first decreases—expected from the standard theory for proximity-induced superconductivity ($\delta R \propto T^{-1/2}$; ref. 19)—then exhibits a minimum around $T_0 = 550$ mK, and increases again at lower temperature. This feature is known as the re-entrance effect^{20,21}. It is expected for a clean NS contact when the barrier resistance of the interface is less than the resistance of the normal sample. It corresponds to the prediction that at $T = 0$ the NS junction has the same conductance as a normal junction²². The minimum of resistance corresponds to $k_B T_0 \approx \hbar/\tau_D$, where τ_D is the time taken for an electron to diffuse from the superconducting interface into the normal reservoir. From $T_0 = 550$ mK, we deduce $\tau_D = 14$ ps. The diffusion time τ_D is rather short, and corresponds to a diffusion length of 500 nm, comparable to the actual size of our sample. Previous measurements of the Fermi distribution function in copper wires²³ have shown that electrons in a $1.5\text{-}\mu\text{m}$ -long wire, with a diffusion time of 350 ps, are in the non-interacting regime for an applied voltage $V = 0.2$ mV. In our sample, τ_D is indeed much smaller than any estimate of τ_{ee} , thus ensuring that even for the largest bias currents the carriers diffusing

through the junction are out of equilibrium. For two-dimensional films of sheet resistance $R_\square$, it is known²⁴ that $\hbar/\tau_{ee} = (e^2/h)k_B T R_\square \ln(\hbar/2e^2 R_\square)$: this gives $\tau_{ee} \approx 100$ ns for $T = 1$ K. For comparison, we note that in 45-nm-thick copper films, a typical energy relaxation time is²³ 1 ns.

In the opposite case of strong electron–electron interaction ($\tau_{ee} \ll \tau_D$), the electrons at local equilibrium produce Johnson–Nyquist noise at an effective temperature higher than the temperature T of the phonons because of the finite thermal conduction to the reservoirs. The heat diffusion process is described by the Wiedeman–Franz law⁹, yielding $F = \sqrt{3/4}$, a value slightly larger than the non-interacting result $F = 1/3$. Even if the carriers in the junction are in the non-interacting regime, or for short enough samples (as in our case), the copper reservoir itself could emit carriers at a higher effective temperature, because the thermalization to the phonon bath happens over large distances¹⁰. The ratio between the sheet resistance of the reservoir and the resistance of the sample controls the temperature of carriers T_{eff} emitted by the normal reservoir into the sample, and therefore the extra Johnson–Nyquist noise. At a large bias voltage, it is known¹⁰ that $T_{\text{eff}} \approx (e/k_B)V[3R_\square \ln(\tau_{ep}/\tau_{ee})/\pi^3 R]^{1/2}$. As the electron–phonon interaction relaxation time τ_{ep} is ~ 200 ns at 1 K (ref. 25), the logarithmic term is approximately unity. Then at $V = 1.27$ mV, and $T = 1.35$ K in our sample (see Fig. 4), T_{eff} would equal 1.9 K. The noise added by this mechanism would be¹⁰ $S_I = (8/3)k_B(T_{\text{eff}} - T)R \approx 0.23 \times 10^{-22} \text{ A}^2 \text{ Hz}^{-1}$ at $V = 1.27$ mV, and is small compared to the observed Andreev-reflection-doubled shot noise (the latter is about 5 times greater).

Figure 3 shows the differential resistance of the sample versus bias current I for $T = 4.2$ and 1.35 K (the base temperature of the experiment). Below $I \approx 2.4$ mA, the niobium electrode is superconducting. Oscillations appear above this critical current that are not discussed here. At $T = 1.35$ K, a dip of about 10% at zero bias develops, which is due to the proximity-induced superconductivity in copper. A resistance plateau appears for $I \approx 1$ mA. Except for the $\sim 10\%$ dip at zero bias currents, the differential resistance does not exhibit any other peculiar features in the range of bias current used for the noise experiment.

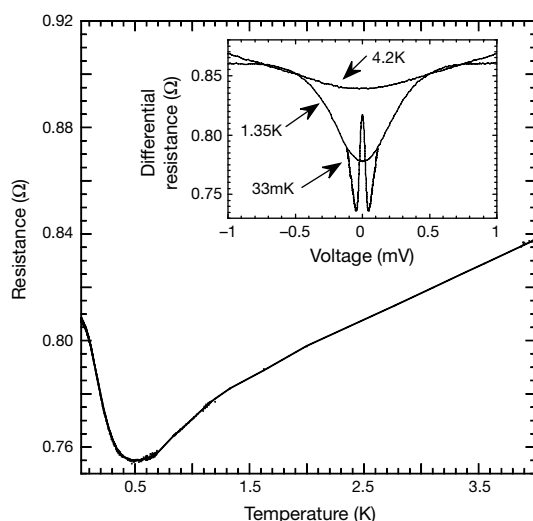


Figure 2 Resistance versus temperature. At $T = 0.5$ K, the resistance reaches a minimum and increases again at lower temperatures. Inset, another signature of the re-entrance effect: at very low temperature (33 mK), a peak appears in the differential resistance at a finite voltage V_0 where $eV_0 \approx k_B T_0$, which is superimposed on the resistance dip due to the proximity effect (see Fig. 3).

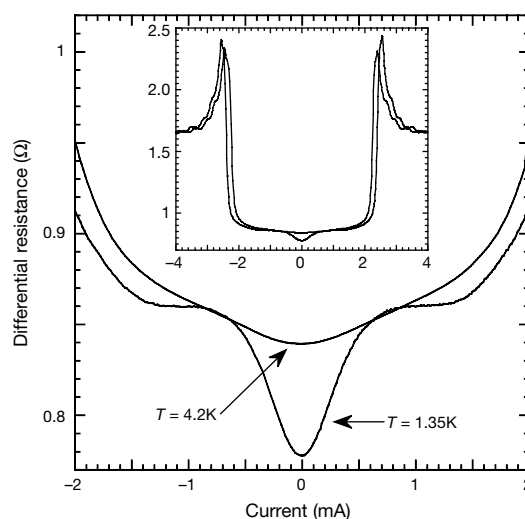


Figure 3 Differential resistance dV/dI as a function of the mean bias current I , at $T = 4.2$ and 1.35 K. When lowering the temperature, the superconductivity induced in copper is enhanced if the bias voltage is not too high: this results in a dip around zero bias on the 1.35 K curve. This proximity-induced superconductivity in copper indicates that the interface between normal metal (Cu) and superconductor (Nb) is excellent. The inset shows the same quantity in the same units, at a larger scale. Above the critical current of ~ 2.4 mA, the Nb film switches into the normal state.

In Fig. 4 we plot the noise power S_I against the bias current. The solid lines show the theoretical dependence given by⁸:

$$S_I = \frac{2}{3} \left[\frac{4k_B T}{R_d} + e^* I \coth \left(\frac{e^* V}{2k_B T} \right) \right] \quad (1)$$

which contains no free parameters apart from the charge of the carriers e^* . The values of I , V and $R_d = dV/dI$ are taken from the data shown in Fig. 3. This formula is strictly established for $e^* = e$, and might in principle change for $e^* = 2e$, but in a single-channel case the exact calculation results only in a small correction²⁶. The substitution of e by $2e$ in equation (1) parallels the treatment of the fractional quantum Hall regime where a similar substitution is justified even if the statistics of the charge-carrying particles differ from fermionic^{1,2}. The statistics of particles could be addressed in a three-terminals branched geometry, where the bosonic type of correlations induced by Andreev reflections are predicted to give positive current–noise correlation^{27–29}.

At $V = 0$, equation (1) gives the usual Johnson–Nyquist noise which does not depend on e^* : $S_I = 4k_B T/R_d$ (at $V = 0$, $R_d = dV/dI = V/I$). At $T = 1.35$ K the agreement with $e^* = 2e$ is excellent up to $I \approx 1.2$ mA ($eV \approx 8k_B T$), including the crossover between the Johnson–Nyquist and the shot-noise regimes occurring for $e^* V = 2k_B T$ (indicated by the arrow in Fig. 4)^{26,30}. At larger voltages, experimental data follow $S_I \approx 2/3 \times 2eI$. The current $I = 1.2$ mA corresponds to a voltage only slightly less than the

superconducting gap (Δ) of bulk niobium ($\Delta/e = 1.35$ mV). Above this value, carriers can enter the niobium electrodes as quasi-particles with a single electron charge. The slope of S_I versus I decreases to about $1/3 \times 2eI$ for the largest voltages that we used.

We do not know of any calculation of the shot noise in SN diffusive junctions at finite voltage and finite temperature. Calculations have been performed^{3,27} of the shot noise in an SN constriction consisting of a hole in a thin screen separating a normal and a superconducting metal at finite voltage and temperature. Unlike the situation in our sample, there is then no distributed disorder scattering in the constriction, but only a barrier of finite transparency ν . We have nevertheless applied the above theory³ to fit our observed voltage dependence of the shot noise ($0 \leq V \leq 2$ mV) at $T = 4.2$ K and 1.35 K, (data not shown). Our best fit is obtained for a superconducting gap of 1 meV and a constriction of resistance $R = 0.8 \Omega = \nu \times R_0 = 0.8 \times 1 \Omega$ (R_0 is the Sharvin resistance of the constriction; that is, its value without any potential barrier). The adjustment gives the correct absolute value for the shot noise within 20% and describes qualitatively the cusp observed near $V \approx \Delta/e$ at $T = 1.35$ K. Better agreement is not expected because our sample is not a SN constriction. In absence of a more relevant theory, equation (1) is the most reliable one at low voltage and low temperature, because it does not rely on any specific microscopic parameters when the SN junction is diffusive. When the voltage is comparable to the superconducting gap, or at a temperature approaching the critical temperature, equation (1) is no longer obeyed, and qualitatively the effective charge is expected to be between $2e$ and e (Fig. 4).

We have shown that the current noise in a short diffusive NS junction at temperatures well below the superconducting critical temperature corresponds to shot noise from carriers with twice the fundamental electron charge. We also note that Andreev-doubled shot noise has recently been observed in concurrent, independent, high-frequency measurements³¹. This doubling of the charge in NS junctions throws light on a subtle aspect of proximity-induced superconductivity: although the zero-temperature conductance of the junction does not depend on whether the emitter is superconducting or not²², the shot noise is doubled when the emitter is superconducting.

To first order, phase coherence seems to be irrelevant to determining the reduction factor of $1/3$ (ref. 11). It is difficult to evaluate whether a normal sample with two normal reservoirs is phase-coherent or not, because coherence only affects the resistance through the small weak localization correction. But with a superconducting reservoir, thanks to the re-entrance effect, the question of phase coherence can be answered. At relatively high temperatures, the electron and the Andreev-reflected hole, whose energies are separated by $k_B T$, lose their relative phase coherence on the timescale $\hbar/k_B T$, suppressing the re-entrance effect. Our results show that the noise doubling persists for temperatures and bias voltages at which the re-entrance effect has disappeared; that is, over a larger energy range than that considered in the theoretical calculation⁴ at zero temperature. □

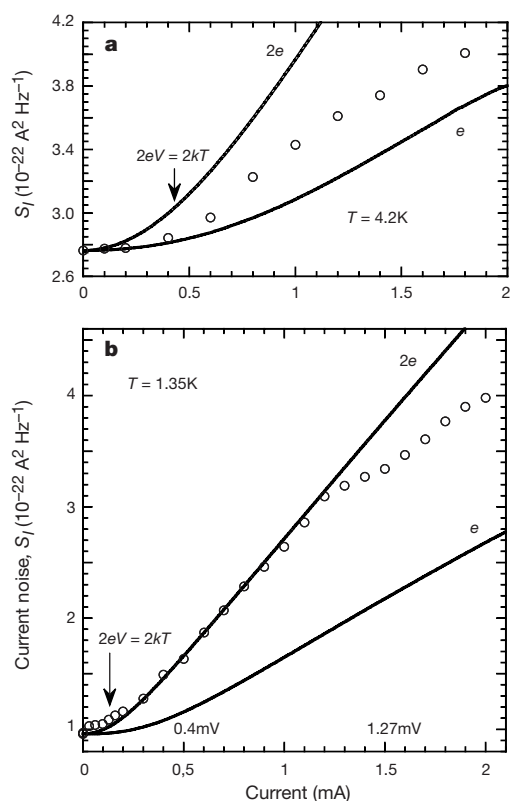


Figure 4 Shot-noise power spectral density as a function of the d.c. bias current I . a, $T = 4.2$ K; b, $T = 1.35$ K. The voltages across the sample are indicated at two values of the current ($I = 0.5$ mA and $I = 1.5$ mA). The data (filled circles) are compared to the predictions (equation (1)) for a normal sample transferring single charges (e) and for the doubled shot noise expected in a short normal-metal/superconductor junction that transfers double charges ($2e$) (see ref. 4). At 1.35 K we observe the predicted doubled shot noise. The discrepancy at higher temperature (4.2 K) and high bias currents could be due to single-charge quasi-particle states that are allowed in the superconductor close to its critical temperature or critical voltage.

Received 9 November 1999; accepted 17 March 2000.

1. Saminadaya, L., Glatli, D. C., Jin, Y. & Etienne, B. Observation of the $e/3$ fractionally charged Laughlin quasiparticle. *Phys. Rev. Lett.* **79**, 2526–2529 (1997).
2. Reznikov, M., de Picciotto, R., Griffiths, T. G., Heiblum, M. & Umansky, V. Observation of quasiparticles with one-fifth of an electron's charge. *Nature* **399**, 238–240 (1999).
3. Khlus, V. A. Current and voltage fluctuations in microjunctions between normal metals and superconductors. *Sov. Phys. JETP* **66**, 1243–1249 (1987).
4. de Jong, M. J. M. & Beenakker, C. W. J. Doubled shot noise in disordered normal-metal-superconductor junctions. *Phys. Rev. B* **49**, 16070–16073 (1994).
5. Lafarge, P., Joyez, P., Esteve, D., Urbina, C. & Devoret, M. H. Two-electron quantization of the charge on a superconductor. *Nature* **365**, 422–424 (1993).
6. Schottky, W. Über spontane Stromschwankungen in verschiedenen Elektrizitätsleitern. *Ann. Phys.* **57**, 541–567 (1918).
7. Beenakker, C. W. J. & Büttiker, M. Suppression of shot noise in metallic diffusive conductors. *Phys. Rev. B* **46**, 1889–1892 (1992).
8. Nagaev, K. E. On the shot noise in dirty metal contacts. *Phys. Lett. A* **169**, 103–107 (1992).

9. Steinbach, A. H., Martinis, J. M. & Devoret, M. H. Observation of hot-electron shot noise in a metallic resistor. *Phys. Rev. Lett.* **76**, 3806–3809 (1996).
10. Henny, M., Oberholzer, S., Strunk, C. & Schönenberger, C. 1/3-shot-noise suppression in diffusive nanowires. *Phys. Rev. B* **59**, 2871–2880 (1999).
11. Blanter, Ya. M. & Büttiker, M. Shot noise in mesoscopic conductors. Preprint condmat/9910158 at (<http://xxx.lanl.gov>) (1999).
12. Andreev, A. F. The thermal conductivity of the intermediate state in superconductors. *Sov. Phys. JETP* **19**, 1228–1231 (1964).
13. Muzykantskii, B. A. & Khmelnitskii, D. E. Quantum shot noise in a normal-metal-superconductor point contact. *Phys. Rev. B* **50**, 3982–3987 (1994).
14. Dieleman, P., Bukkems, H. G., Klapwijk, T. M., Schicke, M. & Gundlach, K. H. Observation of Andreev reflection enhanced shot noise. *Phys. Rev. Lett.* **79**, 3486–3489 (1997).
15. Hoss, T. et al. Multiple Andreev reflection and giant excess noise in diffusive superconductor/normal-metal/superconductor junctions. Preprint cond-mat 9901129 at (<http://xxx.lanl.gov>) (1999).
16. Jehl, X., Payet-Burin, P., Baraduc, C., Calemczuk, R. & Sanquer, M. Andreev reflection enhanced shot noise in mesoscopic SNS junctions. *Phys. Rev. Lett.* **83**, 1660–1663 (1999).
17. Nagaev, K. E. Influence of electron-electron scattering on shot noise in diffusive contacts. *Phys. Rev. B* **52**, 4740–4743 (1995).
18. Jehl, X., Payet-Burin, P., Baraduc, C., Calemczuk, R. & Sanquer, M. Superconducting quantum interference device based resistance bridge for shot noise measurement on low impedance samples. *Rev. Sci. Instrum.* **70**, 2711–2714 (1999).
19. Deutscher, G. & de Gennes, P. G. in *Superconductivity* (ed. Parks, R. D.) 1005–1034 (Dekker, New York, 1969).
20. Courtois, H., Charlat, P., Gandit, Ph., Mailly, D. & Pannetier, B. The spectral conductance of a proximity superconductor and the re-entrance effect. *J. Low Temp. Phys.* **116**, 187–214 (1999).
21. Petrashev, V. T., Shaikhaidarov, R. Sh., Delsing, P. & Claeson, T. Phase-sensitive reentrance into the normal state of mesoscopic SNS structures. *JETP Lett.* **67**, 513–520 (1998).
22. Artemenko, S. N., Volkov, A. F. & Zaitsev, A. V. On the excess current in microbridges S-c-S and S-c-N. *Solid State Commun.* **30**, 771–773 (1979).
23. Pothier, H., Guéron, S., Birge, N. O., Esteve, D. & Devoret, M. H. Energy distribution function of quasiparticles in mesoscopic wires. *Phys. Rev. Lett.* **79**, 3490–3493 (1997).
24. Altshuler, B. L. & Aronov, A. G. in *Electron-Electron Interaction in Disordered Systems* (eds Efros, A. L. & Pollak, M.) 1–153 North-Holland, Amsterdam, 1985).
25. Altshuler, B. L., Aronov, A. G., Gershenson, M. E. & Sharvin, Yu. V. Quantum effects in disordered metal films. *Sov. Sci. A* **9**, 223–354 (1987).
26. Martin, T. Wave packet approach to noise in N-S junctions. *Phys. Lett. A* **220**, 137–142 (1996).
27. Anantram, M. P. & Datta, S. Current fluctuations in mesoscopic systems with Andreev scattering. *Phys. Rev. B* **53**, 16390–16402 (1996).
28. Gramespacher, T. & Büttiker, M. Distribution functions and current-correlations in normal-metal-superconductor hetero-structures. *Phys. Rev. B* **61**, 8125 (2000).
29. Torres, J. & Martin, T. Positive and negative Hanbury-Brown and Twiss correlations in normal metal-superconducting devices. Preprint cond-mat/9906012 at (<http://xxx.lanl.gov>) (1999).
30. Lesovik, B. G., Martin, T. & Torres, J. Josephson frequency singularity in the noise of normal metal-superconductor junctions. *Phys. Rev. B* **60**, 11935–11938 (1999).
31. Kozhevnikov, A. A., Schoelkopf, R. J., Calvet, L. E., Prober, D. E. & Rooks, M. J. Shot noise measurements in diffusive normal-metal-superconductor (N-S) junctions. *J. Low Temp. Phys.* **118**, 671–678 (2000).

Correspondence and requests for materials should be addressed to M.S.
(e-mail: msanquer@cea.fr).

Elastic turbulence in a polymer solution flow

A. Groisman & V. Steinberg

Department of Physics of Complex Systems, Weizmann Institute of Science,
Rehovot 76100, Israel

Turbulence is a ubiquitous phenomenon that is not fully understood. It is known that the flow of a simple, newtonian fluid is likely to be turbulent when the Reynolds number is large (typically when the velocity is high, the viscosity is low and the size of the tank is large^{1,2}). In contrast, viscoelastic fluids³ such as solutions of flexible long-chain polymers have nonlinear mechanical properties and therefore may be expected to behave differently. Here we observe experimentally that the flow of a sufficiently elastic polymer solution can become irregular even at low velocity, high viscosity and in a small tank. The fluid motion is excited in a broad range of spatial and temporal scales, and we observe an increase in the flow resistance by a factor of about twenty. Although the Reynolds number may be arbitrarily low, the observed flow has all the main features of developed turbulence. A

comparable state of turbulent flow for a newtonian fluid in a pipe would have a Reynolds number as high as 10^5 (refs 1, 2). The low Reynolds number or 'elastic' turbulence that we observe is accompanied by significant stretching of the polymer molecules, resulting in an increase in the elastic stresses of up to two orders of magnitude.

Motion of simple, low molecular weight, newtonian fluids is governed by the Navier–Stokes equation^{1,2}. This equation has a nonlinear term, which is due to fluid inertia. The ratio between the nonlinearity and viscous dissipation is given by the Reynolds number, $Re = VL/\nu$, where V is velocity, L is characteristic size and ν is kinematic viscosity of the fluid. When Re is high, nonlinear effects are strong and the flow is likely to be turbulent; therefore, turbulence is a paradigm for a strongly nonlinear phenomenon^{1,2}.

Solutions of flexible high molecular weight polymers differ from newtonian fluids in many aspects³. The most notable elastic property of the polymer solutions is that stress does not immediately become zero when the fluid motion stops, but rather decays with some characteristic time, λ , which can reach seconds and even minutes. The equation of motion for dilute polymer solutions differs from the Navier–Stokes equation by an additional linear term arising from the elastic stress, τ (ref. 3). Because the elastic stress is caused by stretching of the polymer coils, it depends on history of motion and deformation of fluid elements along their flow trajectories. This implies a nonlinear relationship between τ and the rate of deformation in a flow³. The nonlinear mechanical properties of polymer solutions are well manifested in their large extensional viscosity at high rates of extension⁴ and in the Weissenberg effect^{3,5}. The degree of nonlinearity in the mechanical properties is expressed by the Weissenberg number, $Wi = V\lambda/L$, which is a product of characteristic rate of deformation and the relaxation time, λ .

We considered whether the nonlinearity of mechanical properties of a fluid can give rise to turbulent flow when the equation of motion is linear. For a polymer solution this corresponds to a state in which the Weissenberg number is large, while the Reynolds number is small. This situation can be realized if the parameter of

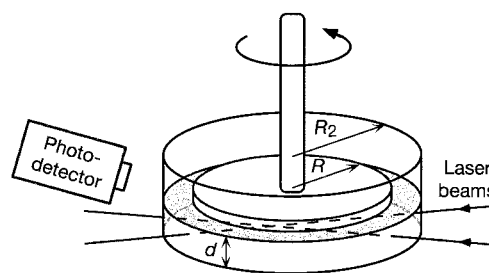


Figure 1 The experimental set-up. A stationary cylindrical cup with a plain bottom (the lower plate) is concentric with the rotating upper plate, which is attached to the shaft of a commercial rheometer. The radii of the upper and the lower plates are $R = 38$ mm and $R_2 = 43.6$ mm, respectively. The liquid is filled until a level d of 10 mm unless otherwise stated. The upper plate just touches the surface of the liquid. A special cover is used to minimize evaporation of the liquid. We used a solution of 65% saccharose and 1% NaCl in water, viscosity $\eta_s = 0.324$ Pa s, as a solvent for the polymer. We added polyacrylamide ($M_w = 18,000,000$; Polysciences) at a concentration of 80 p.p.m. by weight. The solution viscosity was $\eta = 0.424$ Pa s at $\dot{\gamma} = 1$ s⁻¹. The relaxation time, λ , estimated from the phase shift between the stress and the shear rate in oscillatory tests, was 3.4 s. The temperature is stabilized at 12 °C by circulating water under the steel lower plate. The walls of the cup are transparent which allows Doppler velocimeter measurements by collecting light scattered from the crossing point of two horizontal laser beams. In experiments where the flow has to be viewed from below, the lower plate is made from plexiglass and a mirror tilted by 45° is placed under the lower plate. The flow patterns are then captured by a CCD camera at the side and the temperature is stabilized by circulating air in a closed box.

elasticity, $Wi/Re = \lambda \nu / L^2$, is large enough. An important observation concerning the influence of the nonlinear mechanical properties on flow was made about a decade ago, when purely elastic instability was experimentally identified in curvilinear shear flows^{6,7}. This instability occurs at moderate Wi and vanishingly small Re and is driven by the elastic stresses^{7,9}. As a result of the instability, secondary and in general oscillatory vortex flows develop, and flow resistance increases^{6–10}. Flow instabilities in elastic liquids are reviewed in refs 11, 12.

There is no commonly accepted unique definition of turbulent flow², and it is usually identified by its main features^{1,2}. Turbulence implies fluid motion in a broad range of spatial and temporal scales, so that many degrees of freedom are excited in the system. A practically important characteristic of turbulent flows is a large increase in the flow resistance compared with an imaginary laminar flow with the same Re . We show how these main features of turbulence appear in a flow of a highly elastic polymer solution at low Reynolds numbers.

For our experiments we chose a swirling flow between two parallel disks (Fig. 1), and a dilute solution of high molecular weight polyacrylamide in a viscous sugar syrup as the working fluid. The curvature ratio was made quite high, $d/R = 0.263$, to provide

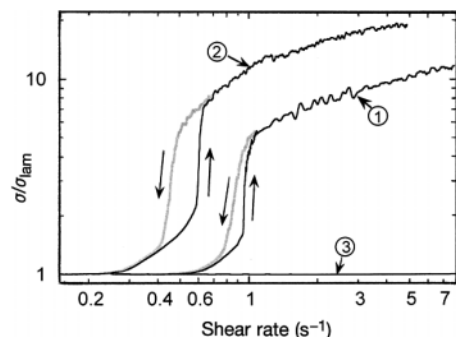


Figure 2 Stress ratio versus shear rate. We plot the ratio of the average stress, σ , measured in the flow, to the stress, σ_{lam} , in the laminar flow with the same boundary conditions as a function of the shear rate, $\dot{\gamma}$. The curves 1 and 2 are for the polymer solution flow with $d = 10$ mm and 20 mm, respectively. The shear rate was gradually varied in time, very slowly (by about $10\% \text{ h}^{-1}$) in the transition region, and faster below and above it. Thin black lines represent increasing $\dot{\gamma}$; thick grey lines represent decreasing $\dot{\gamma}$. Curve 3 represents the pure solvent. Mechanical degradation of the polymers was quite small at shear rates below 1.5 s^{-1} and 1 s^{-1} for $d = 10$ mm and 20 mm, respectively. The dependences of $\sigma/\sigma_{\text{lam}}$ on $\dot{\gamma}$ in those regions were therefore reproducible in consecutive runs within about 1%. Degradation effects became appreciable at higher shear rates, and elasticity typically decreased by up to 10% as a result of the runs shown by the curves 1 and 2.

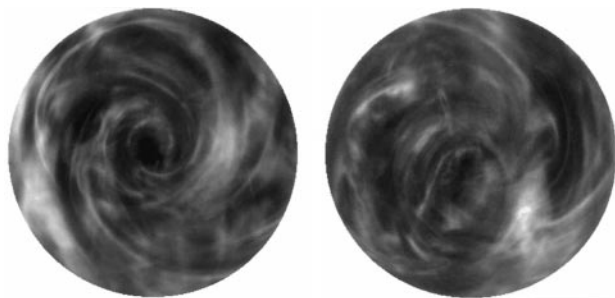


Figure 3 Two snapshots of the flow at $Wi = 13$, $Re = 0.7$. The flow under the black upper plate is visualized by seeding the fluid with light reflecting flakes (1% of the Kalliroscope liquid). The fluid is illuminated by ambient light. Although the pattern is quite irregular, structures that appear tend to have spiral-like forms. The dark spot in the middle corresponds to the centre of a big persistent thoroidal vortex that has dimensions of the whole set-up.

destabilization of the primary shear flow and development of the secondary vortical fluid motion at lower shear rates^{7,10}. (The flow between two plates with small d/R has been studied previously in the context of the purely elastic instability¹⁰.) We mounted the whole flow set-up on top of a commercial viscometer (AR-1000 of TA-instruments) to measure precisely the angular velocity, ω , of the rotating upper plate and the torque applied to it. In this way we were able to estimate the average shear stress, σ , in the polymer solution and to compare it with the stress in the laminar flow, σ_{lam} , with the same applied shear rate. In newtonian fluids the ratio $\sigma/\sigma_{\text{lam}}$ generally grows with Re as the flow becomes increasingly irregular, and the magnitude of $\sigma/\sigma_{\text{lam}}$ can be considered as a measure of strength of turbulence and turbulent resistance. In our set-up σ becomes 30% higher than σ_{lam} at $Re = 70$, which can be regarded as a point when inertial effects become significant.

The dependence of $\sigma/\sigma_{\text{lam}}$ on the shear rate, $\dot{\gamma} = \omega R/d$, for flow of the polymer solution in the experimental system is shown in Fig. 2 (first curve). At a value of $\dot{\gamma}$ of about 1 s^{-1} (corresponding to $Wi \equiv \lambda \dot{\gamma} = 3.5$), a sharp transition occurs that appears as a significant increase in the apparent viscosity. The Reynolds number at the transition point is about 0.3, which means that the inertial effects are quite negligible. The transition has pronounced hysteresis,

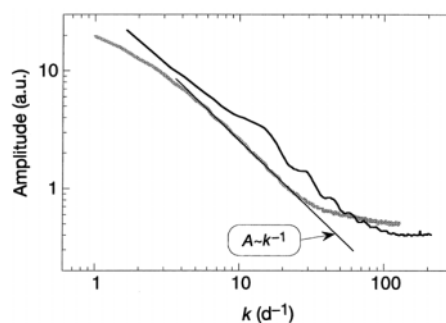


Figure 4 Average Fourier spectra of the brightness profiles. These are taken along the diameter (thin black line) and along the circumference at a radius of $2d$ (thick grey line). We averaged over a long series of flow pattern snapshots taken in consecutive moments of time. The wavelength is measured in units of d , so that the wavenumber, k , of unity corresponds to a length of $2\pi d$. The spectrum taken along the diameter apparently differs from the azimuthal spectrum by a series of broad peaks. This may be a manifestation of the fact that the flow is not completely structureless and homogeneous along the radial direction (see Fig. 3). The visualization method that we used (Fig. 3) does not provide direct information about the fluid velocity; therefore, the exact value of the exponent in the power law fit, $A \approx K^{-1}$, has no special meaning.

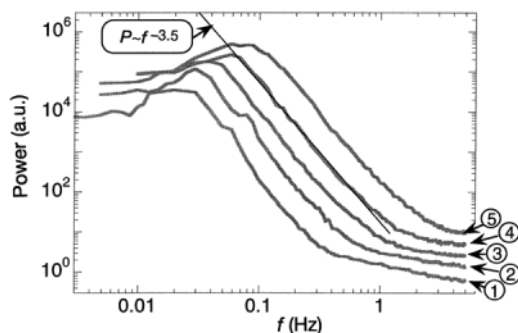


Figure 5 Power spectra of velocity fluctuations. The data are obtained at different rates, $\dot{\gamma}$ in the standard set-up. The fluid velocity was measured by a laser Doppler velocimeter in the centre of the flow. Curves 1–5 correspond to $\dot{\gamma} = 1.25, 1.85, 2.7, 4$ and 5.9 s^{-1} , respectively (all above the transition point $\dot{\gamma} \approx 1$, Fig. 2). The power, P , of fluctuations is fitted by a power law, $P \approx f^{-3.5}$, for $\dot{\gamma} = 4 \text{ s}^{-1}$ over about a decade in frequencies, f .

which is typical for the purely elastic flow instability⁹. The ratio $\sigma/\sigma_{\text{lam}}$ keeps growing with the shear rate, and at the highest $\dot{\gamma}$ that has been reached the flow resistance is about 12 times larger than is found in laminar flow. In the same range of shear rates, flow of the pure solvent is completely laminar and the ratio $\sigma/\sigma_{\text{lam}}$ is unity within the resolution of the viscometer (about 1%). To make sure that the observed flow phenomena were indeed caused by the solution elasticity, we measured σ for a few solutions that had the same polymer concentration but different relaxation times, λ . The curves of $\sigma/\sigma_{\text{lam}}$ coincided, when plotted against Wi , whereas the Reynolds number turned out to be completely irrelevant (see also ref. 9). The growth of the resistance in the polymer solution flow becomes even larger when the size of the gap is increased (Fig. 2, second curve). At this point the ratio $\sigma/\sigma_{\text{lam}}$ reaches a value of 19. Such growth of the flow resistance is found for newtonian fluids in the same flow geometry at Re of about 2×10^4 . For flow in a circular pipe this value of $\sigma/\sigma_{\text{lam}}$ is reached at $Re \approx 10^5$, which is usually considered as a region of rather developed turbulence¹.

Two representative snapshots of the polymer solution flow above the transition (at $\dot{\gamma} = 4 \text{ s}^{-1}$) are shown in Fig. 3. The flow patterns are very irregular and structures of different sizes appear. This visual impression is confirmed by a more careful analysis. Average Fourier spectra of the brightness profiles along the diameter and along the circumference exhibit power law decay over a decade in the wavenumber domain (Fig. 4).

Characteristic time spectra of velocity fluctuations at different shear rates are shown in Fig. 5. Flow velocity was measured in the horizontal plane in the centre of the set-up, where its average value is zero. As the shear rate is raised the power of fluctuations increases and characteristic frequencies become higher, but the general form of the spectra remains very much the same. In particular, just as for the spatial spectra in Fig. 4, there is a region of a power law decay, which spans about a decade in frequencies. This power law dependence in the broad ranges of spatial and temporal frequencies actually means that the fluid motion is excited at all those spatial and temporal scales. Spectra of radial and azimuthal velocities taken at different points with non-zero average flow had the same general appearance and close values of exponents in the power law decay range.

In summary, we conclude that the flow of the elastic polymer solution at sufficiently high Wi has all the main features of developed turbulence. By the strength of the turbulent resistance, and by the span of scales in space and time, where the fluid motion is excited, the observed flow can be compared with turbulence in a newtonian fluid in a pipe at Re of about 10^5 . This apparently turbulent flow arises solely because of the nonlinear mechanical properties of the elastic polymer solution. We therefore call the phenomenon elastic turbulence, in contrast to the usual inertial turbulence which is observed in newtonian fluids at high Re . (The name 'elastic turbulence' has been used before for designation of apparently disordered flows in polymeric liquids¹³. No attempt has been ever made, however, to characterize those flows quantitatively.)

Elastic turbulence has many features that are in sharp contradiction to the concepts of newtonian fluid mechanics. On the one hand, velocity required for excitation of inertial turbulence in a newtonian fluid is proportional to the fluid viscosity. On the other hand, the polymer relaxation time, λ , usually grows proportionally to the viscosity. Because at constant d/R the transition to elastic turbulence occurs at a certain value of the Weissenberg number, $Wi = \lambda \dot{\gamma}$, one can excite turbulence at lower velocities by choosing more viscous polymer solutions. Indeed, using a solution of polymers in a very viscous sugar syrup, we observed transition to the elastic turbulence at a rotation rate of 0.05 s^{-1} (corresponding to $Re \approx 10^{-3}$). Further, in an elastic polymer solution, the scale of time, λ , does not depend on the size of the system; therefore as long as the ratio d/R is preserved, transition to turbulence should occur at the same ω , and the dependence of $\sigma/\sigma_{\text{lam}}$ on $\dot{\gamma}$ should not change

with the size of the system. We repeated the measurements of σ in a small set-up having all the dimensions reduced by a factor of 4 as compared with the standard system. The dependence of $\sigma/\sigma_{\text{lam}}$ on $\dot{\gamma}$ was found to be the same (data not shown) as in Fig. 2, whereas the characteristic velocities and the Reynolds numbers were lower by factors of 4 and 16, respectively. We therefore believe that by using polymer solutions with sufficiently high elasticity we can excite turbulent motion at arbitrary low velocities and in arbitrary small tanks. (The size of the tank still has to be large compared with the size of the polymer coils.)

An important question about the elastic turbulence is where the turbulent resistance comes from. In the inertial turbulence the origin of the large resistance is the Reynolds stress, which is connected with high kinetic energy of the turbulent motion and takes a major part in the momentum transfer in the flow. Elastic turbulence occurs at low Reynolds numbers. From our velocity measurements in the standard set-up, contribution of the Reynolds stress to the flow resistance could be estimated as being less than 0.5%. The contribution of the viscous shear stress of the newtonian solvent, averaged across the fluid layer, is always the same as in laminar flow and cannot change. Thus, the whole increase in the flow resistance should be due to the elastic stress. The data shown in Fig. 2 (second curve) imply that the polymer contribution to the stress increases by a factor of up to 65, as compared with laminar flow with the same average shear rate. This suggestion agrees very well with our measurements of relaxation of the shear stress after the fluid motion is stopped. The elastic part, τ , of the whole stress is identified by its slow relaxation with a characteristic time of the order λ . In elastic turbulence this slowly relaxing part can become two orders of magnitude larger than in laminar flow with the same shear rate. This major growth of the elastic stress should be connected with vast extension of the polymer molecules in the turbulent flow.

Thus, elastic turbulence apparently develops as follows. The polymer molecules are stretched in the primary shear flow, which makes it unstable and causes irregular secondary flow. This flow acts back on the polymer molecules, stretching them further and becoming increasingly turbulent, until a kind of saturated dynamic state is reached. The density of the elastic energy of the stretched polymers can be estimated as $Wi \tau/2$, and should therefore increase in the elastic turbulent flow by about the same factor as the elastic stress, τ , while the kinetic energy remains small. □

Received 6 December 1999; accepted 2 March 2000.

- Landau, L. D. & Lifschitz, E. M. *Fluid Mechanics* (Pergamon, Oxford, 1987).
- Tritton, D. J. *Physical Fluid Dynamics* (Clarendon, Oxford, 1988).
- Bird, R. B., Curtiss, C. F., Armstrong, R. C. & Hassager, O. *Dynamics of Polymeric Liquids* (Wiley, New York, 1987).
- Tirtaatmadja, V. & Sridhar, T. A filament stretching device for measurement of extensional viscosity. *J. Rheology* **37**, 1081–1102 (1993).
- Weissenberg, K. A continuum theory of rheological phenomena. *Nature* **159**, 310–311 (1947).
- Magda, J. J. & Larson, R. G. A transition occurring in ideal elastic liquids during shear flow. *J. Non-Newtonian Fluid Mech.* **30**, 1–19 (1988).
- Muller, S. J., Larson, R. G. & Shaqfeh, E. S. G. A purely elastic transition in Taylor–Couette flow. *Rheol. Acta* **28**, 499–503 (1989).
- Larson, R. G., Shaqfeh, E. S. G. & Muller, S. J. A purely viscoelastic instability in Taylor–Couette flow. *J. Fluid Mech.* **218**, 573–600 (1990).
- Groisman, A. & Steinberg, V. Mechanism of elastic instability in Couette flow of polymer solutions: experiment. *Phys. Fluids* **10**, 2451–2463 (1998).
- Byars, J. A., Öztekin, A., Brown, R. A. & McKinley, G. H. Spiral instabilities in the flow of highly elastic fluids between rotating parallel disks. *J. Fluid Mech.* **271**, 173–218 (1994).
- Larson, R. G. Instabilities in viscoelastic flows. *Rheol. Acta* **31**, 213–263 (1992).
- Shaqfeh, E. S. G. Purely elastic instabilities in viscometric flows. *Annu. Rev. Fluid. Mech.* **28**, 129–185 (1996).
- Giesekus, H. W. Non-linear effects in the flow of visco-elastic fluids through slits and holes. *Rheol. Acta* **7**, 127–138 (1968).

Acknowledgements

This work was supported by the Minerva Center for Nonlinear Physics of Complex Systems.

Correspondence and requests for materials should be addressed to A. G.

Rapid prototyping of patterned functional nanostructures

Hongyou Fan[‡], Yunfeng Lu[†], Aaron Stump[‡], Scott T. Reed[†], Tom Baer[†], Randy Schunk[†], Victor Perez-Luna[‡], Gabriel P. López[‡] & C. Jeffrey Brinker^{†‡}

[‡] The University of New Mexico Center for Micro-Engineered Materials and Department of Chemical and Nuclear Engineering, Albuquerque, New Mexico 87131, USA

[†] Sandia National Laboratories, Albuquerque, New Mexico 87185, USA

Living systems exhibit form and function on multiple length scales and at multiple locations. In order to mimic such natural structures, it is necessary to develop efficient strategies for assembling hierarchical materials. Conventional photolithography, although ubiquitous in the fabrication of microelectronics and microelectromechanical systems, is impractical for defining feature sizes below 0.1 micrometres and poorly suited to pattern chemical functionality. Recently, so-called 'soft' lithographic approaches¹ have been combined with surfactant^{2,3} and particulate⁴ templating procedures to create materials with multiple levels of structural order. But the materials thus formed have been limited primarily to oxides with no specific functionality, and the associated processing times have ranged from hours to days. Here, using a self-assembling 'ink', we combine silica-surfactant self-assembly with three rapid printing procedures—pen lithography, ink-jet printing, and dip-coating of patterned self-assembled monolayers—to form functional, hierarchically organized structures in seconds. The rapid-prototyping procedures we describe are simple, employ readily available equipment, and provide a link between computer-aided design and self-assembled nanostructures. We expect that the ability to form arbitrary functional designs on arbitrary surfaces will be of practical importance for directly writing sensor arrays and fluidic or photonic systems.

Since the discovery of surfactant-templated silica mesophases⁵, considerable effort has been devoted to the development of molecular-scale, organic modification schemes to impart useful functionality to the pore surfaces^{6–13}. Concurrently a variety of patterning strategies have been developed to define macroscopically the shapes of deposited thin-film mesophases and their locations on the substrate surface^{2–4}. Our approach combines molecular-scale, evaporation-induced self-assembly (EISA)¹⁴ of organically modified mesophases^{6–13} with macroscopic, evaporative printing procedures. We report the rapid fabrication of hierarchical structures exhibiting

form and function on multiple length scales and at multiple locations. At the molecular scale, functional organic moieties (Table 1) are positioned on pore surfaces; on the mesoscale, mono-sized pores are organized into one-, two- or three-dimensional networks, providing size-selective accessibility from the gas or liquid phase; and on the macroscale, two-dimensional (2D) arrays and fluidic or photonic systems are defined.

Vital to rapid patterning procedures is the use of stable, homogeneous inks that on evaporation undergo self-assembly to form the desired organically modified silica-surfactant mesophase. For this purpose we prepared oligomeric silica sols in ethanol/water solvents at a hydronium ion concentration ($[H_3O^+] \approx 0.01$ M) designed to minimize the siloxane condensation rate, thereby enabling facile silica-surfactant self-assembly during the brief time span of the writing operation (several seconds). Surfactants were added at an initial concentration c_0 , much less than the critical micelle concentration, c.m.c., ensuring homogeneity and longevity of the ink.

As a pattern of ink is metered onto a surface, preferential evaporation of ethanol causes enrichment of water, surfactant and silicates, establishing a complex three-dimensional (3D; longitudinal and radial) gradient in their respective concentrations (see, for example, the 3D finite-element simulation of pen lithography in Fig. 1b). Where the c.m.c. is exceeded, cooperative silica-surfactant self-assembly creates micelles. Further evaporation, predominantly of water, promotes the continuous self-organization of micelles into silica-surfactant liquid-crystalline mesophases. Incipient liquid-crystalline domains are nucleated at liquid-vapour interfaces¹⁵ (at concentration $c < \text{c.m.c.}$) and grow inward along compositional trajectories established by the steep, 3D evaporation-induced concentration gradient (Fig. 1). The amphiphilic nature of some organosilanes like tridecafluoro-1,1,2,2-tetrahydrooctyltriethoxysilane (TFTS; compound **1** in Table 1) causes them to behave as co-surfactants; this positions the hydrophilic $Si(OR)_2(OH)_{3-x}$ head-groups in close proximity to the silica oligomers where they are incorporated into the framework upon further hydrolysis and condensation, thereby localizing covalently attached R' moieties on the pore surfaces. Hydrophobic but alcohol-soluble organic molecules like rhodamine-B partition into the hydrophobic micellar interiors upon ethanol evaporation¹⁶, and ultimately become compartmentalized within the channel network of the resulting mesophase. Retention of the covalently attached functional organic moieties after surfactant removal by pyrolysis was confirmed using ²⁹Si and ¹³C magic-angle spinning (MAS) NMR spectroscopy (see ²⁹Si MAS NMR spectra and corresponding thermogravimetric data in Supplementary Information Figs 1–3). Fluorescence imaging and ultraviolet-visible spectroscopy were used to confirm retention and functionality of optically active ligands and molecules.

Table 1 Functional organosilanes and properties of resultant thin-film mesophases

	Functional silanes/additives† $R'-Si(OR)_3$	Mesophase	Pore size* (Å)	Surface area* (m ² g ⁻¹)	Properties and applications
1	F₃C(CF₂)₈CH₂CH₂Si(OC₂H₅)₃ Tridecafluoro-1,1,2,2-tetrahydrooctyltriethoxysilane (TFTS)	3-dH	25	850	Hydrophobic; low- k dielectrics
2	HS-(CH₂)₃Si(OCH₃)₃ Mercaptopropyltrimethoxysilane (MPS)	3-dH	25	1,060	Coupling of noble metals
3	NH₂-(CH₂)₃Si(OCH₃)₃ Aminopropyltrimethoxysilane (APS)	Cubic	22	750	Coupling of noble metals, dye, and bioactive molecules
4	Dye §-NH-(CH ₂) ₃ Si(OCH ₃) ₃	Cubic	21	545	pH sensitive
5	O₂N -C ₆ H ₄ (NO ₂)-NH(CH ₂) ₃ Si(OC ₂ H ₅) ₃	3-dH	22	560	Chromophore; nonlinear optical material ($\chi^{(2)}$)
6	(H ₃ C ₂ O) ₃ Si CH₂CH₂Si(OC₂H₅)₃	Cubic	40	430	Low- k dielectrics

* Pore size and surface area were determined from N₂ sorption isotherms obtained at -196°C, using a surface acoustic wave (SAW) technique²². Mass change due to nitrogen sorption was monitored (~80 pg cm⁻² sensitivity) as a function of nitrogen relative pressure. Pore size and surface area were determined from the isotherms using the BET equation and the BJH algorithm, respectively.

† Functional groups are retained through selective surfactant removal during heat treatment in nitrogen. TGA and DTA were used to establish the appropriate temperature window enabling complete surfactant removal without silane decomposition.

‡ Additives investigated include rhodamine-B, cytochrome c (Fluka), oil blue N, disperse yellow 3 (Aldrich), silver ions and silver nanoparticles.

§ **4** was prepared by a conjugation reaction between a thin-film mesophase containing **3** and the dye molecule (5,6-carboxyfluorescein, succinimidyl ester (5,6-FAM, SE) from Molecular Probes).

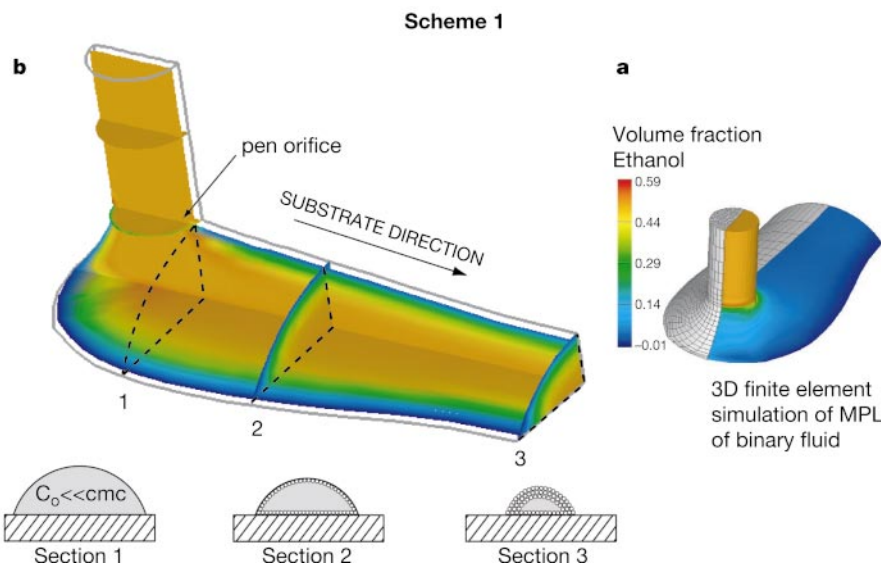


Figure 1 Scheme 1: micropen lithography (MPL) of a surfactant-templated mesophase. **a**, Simulation of 3D, binary fluid pattern dispensed on a flat substrate with the following parameters: pen orifice, 50.0 μm ; substrate speed, 2.5 cm s^{-1} ; and fluid injection rate (inlet velocity), 4.0 cm s^{-1} . Colour contours represent evaporation-induced, 3D gradients in alcohol composition. Residence times for fluid elements entering at the pen orifice and leaving at section 3 ranged from 0.23 to 0.30 ms. Fluid was modelled as 54 vol.% ethanol and 46 vol.% non-volatile phase with Reynolds number $\text{Re} = 1.25$ and $\text{Ca} = 0.000833$. An *ad hoc* value of 45° was chosen for the static contact angle. Note that this angle persists at all points on the dynamic contact line because of the dominance of surface

tension at this low value of Ca . **b**, The initially homogeneous sol metered on to the moving substrate experiences preferential evaporation of alcohol creating a complex 3D (longitudinal and radial) gradient in the concentrations of water and non-volatile surfactant and silicate species. Progressive enrichment of silica and surfactant induces micelle formation and subsequent growth of silica-surfactant mesophases inward from the liquid-vapour interface as recently demonstrated for aerosols²³. The numerical method utilized for **a** and **b** consisted of a 3D finite-element discretization of the Navier-Stokes equations augmented with a 3D boundary-fitted mesh motion algorithm to track the free surface^{30,31}. Special relations at the 3D dynamic wetting line were also applied.

Scheme 1 (Fig. 1) schematically illustrates direct writing of a mesoscopically ordered nanostructure, using micropen lithography, MPL¹⁷. Figure 2a shows a macroscopic pattern formed in several seconds by MPL of a rhodamine-B-containing solution on a hydrophilic surface. The inset in Fig. 2a shows the corresponding fluorescence image of several adjacent stripes acquired through a 610-nm bandpass filter, demonstrating retention of rhodamine-B functionality, and the transmission electron microscopy (TEM) image (Fig. 2b) reveals the ordered pore structure characteristic of a cubic thin-film mesophase. The MPL line width can vary from micrometres to millimetres. It depends on such factors as pen dimension¹⁸, wetting characteristics, evaporation rate, capillary

number ($\text{Ca} = \text{ink viscosity} \times \text{substrate speed/surface tension}$) and ratio of the rates of ink supply and withdrawal (inlet velocity/substrate velocity). The effect of wetting has been demonstrated by performing MPL on substrates prepatterned with hydrophobic, hydrophilic or mixed self-assembled monolayers (SAMs). Generally,

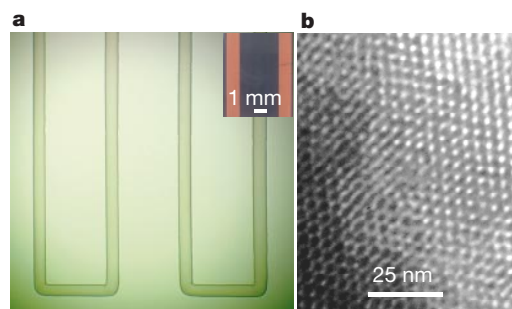


Figure 2 Meandering patterned mesophase created by MPL. **a**, Optical micrograph of patterned rhodamine-B-containing silica mesophase deposited on an oxidized [100]-oriented silicon substrate at a speed of 2.54 cm s^{-1} . Inset is a fluorescence image of rhodamine-B emission acquired through a 610-nm bandpass filter, demonstrating retention of rhodamine-B functionality. **b**, Representative TEM micrograph of a fragment of the patterned rhodamine-B-containing film corresponding to a [110]-oriented cubic mesophase with lattice constant $a = 10.3 \text{ nm}$. The sol was prepared by adding 0.01 wt% rhodamine-B to a silica/4wt% Brij-56 sol. The TEOS:EtOH:water:HCl:Brij-56:rhodamine-B molar ratio = 1:22:5.0:0.004:0.075:2.6 $\times 10^{-5}$.

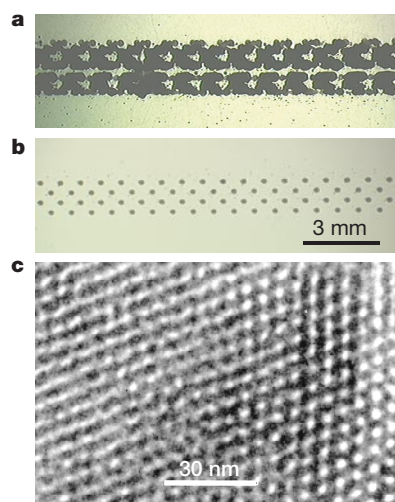


Figure 3 Patterned dot arrays created by ink-jet printing, IJP. **a**, Optical micrograph of a dot array created by IJP of standard ink (Hewlett-Packard Co.) on a non-adsorbent surface. **b**, Optical micrograph of an array of hydrophobic, mesoporous silica dots created by evaporation-induced silica-surfactant self-assembly during IJP on an oxidized [100]-oriented silicon substrate followed by calcination. **c**, Representative TEM micrograph of a dot fragment prepared as in **b**. The sol was prepared with molar ratio TEOS:TFTS:EtOH:water:HCl:Brij-56 = 1:0.05:22.0:5.0:0.004:0.075. The dot pattern used in **a** and **b** was designed using Microsoft PowerPoint 98 software. The printing rate was approximately 80 dots s^{-1} and printer resolution 300 dots per inch. The resolution achieved compared to standard ink and our ability to selectively functionalize the ink suggest applications in display technologies.

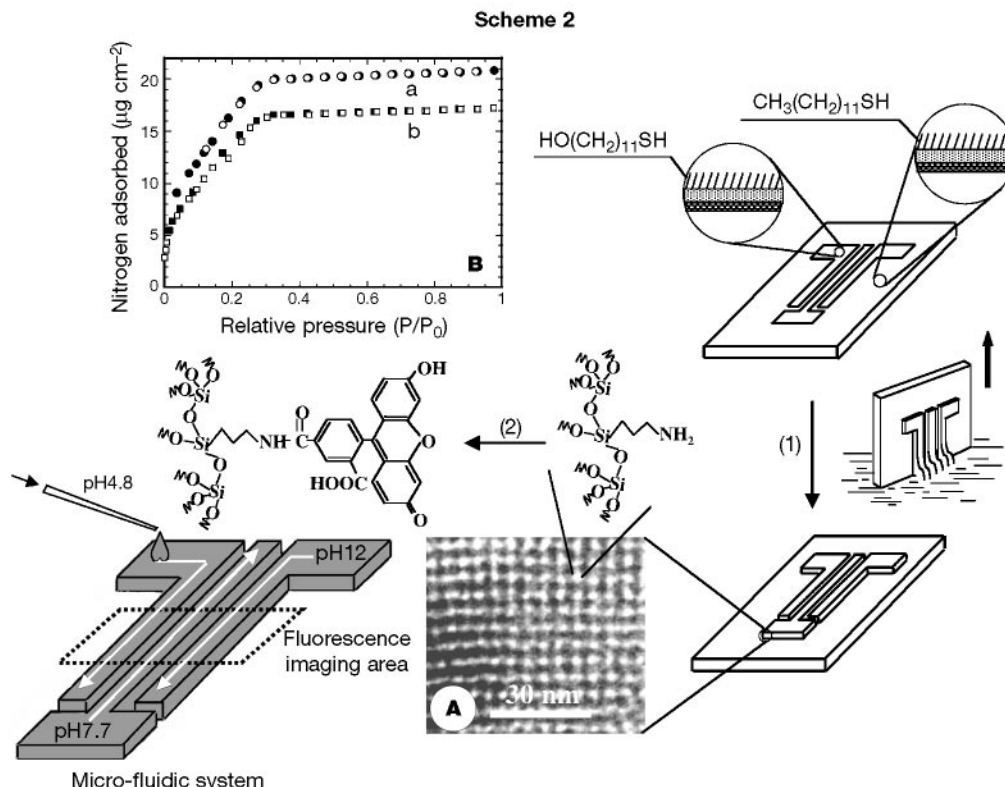


Figure 4 Scheme 2: patterned functional mesostructure formed by selective de-wetting. Using microcontact printing or electrochemical desorption techniques, substrates are prepared with patterns of hydrophilic, hydroxyl-terminated SAMs and hydrophobic methyl-terminated SAMs. Preferential ethanol evaporation during dip-coating (1), causes water enrichment and selective dewetting of the hydrophobic SAMs. Correspondingly film deposition occurs exclusively on the patterned hydrophilic SAMs. The sol was prepared by adding aminopropyltrimethoxysilane ($\text{NH}_2(\text{CH}_2)_3\text{Si}(\text{OCH}_3)_3$, APS) to a silica/4wt% Brij-56 sol, resulting in a final molar ratio TEOS:APS:EtOH:water:HCl:Brij-56 = 1:0.8:22:5.0:0.011:0.075. Selective dewetting followed by calcination results in a patterned, amine-functionalized, cubic mesoporous film as is evident from the plan-view TEM micrograph (inset A), showing a [100]-oriented cubic mesophase with $a = 10.3 \text{ nm}$

line widths are reduced by increasing the contact angle and by reducing the pen orifice dimension and inlet/substrate velocity ratio.

The advantages of MPL are that we can use any desired combination of surfactant and functional silane as ink to print selectively different functionalities at different locations. Furthermore, we can use computer-aided design (CAD) to define arbitrary 2D patterns that can be written on arbitrary surfaces. For example, we have demonstrated writing rhodamine-containing mesophases (refractive index $n = 1.2\text{--}1.3$) on aerogel¹⁹ and emulsion-templated thin films ($n = 1.03\text{--}1.10$), thereby directly defining optical waveguide structures potentially useful for lasing²⁰.

MPL is best suited to write continuous patterns. Patterned macroscopic arrays of discrete mesostructures can be obtained readily by combining EISA with aerosol processing schemes like ink-jet printing, IJP^{21,22}. Figure 3b shows an optical micrograph of an array of hydrophobic, mesoporous spots formed on a silicon substrate by IJP of a TFS-modified ink. The IJP process dispenses the ink (prepared as for MPL) as monosized, spherical aerosol droplets. On striking the surface, the droplets adopt a new shape that balances surface and interfacial energies. Accompanying evaporation creates within each droplet a gradient in surfactant concentration that drives radially directed silica-surfactant self-assembly inward from the liquid-vapour interface²³. The TEM micrograph (Fig. 3c) shows the ordered mesoporosity of a calcined, fluoroalkylated silica mesophase formed by IJP. The link to

and nitrogen adsorption-desorption isotherm (inset B, curve a) acquired for the thin-film specimen using a surface acoustic wave³² (SAW) technique. The dye conjugation reaction (2) was conducted by immersion in a 0.00002 mM solution of 5,6-FAM, SE (Table 1) prepared in dimethylsulphoxide (DMSO) followed by exhaustive, successive washing in DMSO, ethanol and water. The SAW-based nitrogen adsorption-desorption isotherm of the dye-conjugated mesoporous film is shown in inset B, curve b, confirming its pore accessibility. BET analyses of the sorption isotherms indicate that the dye conjugation reaction reduces the surface area from 750 to $545 \text{ m}^2 \text{ g}^{-1}$ and the hydraulic radius from 2.2 to 2.1 nm , but pore accessibility is completely retained as evident from combined TEM, SAW and fluorescence-imaging results (Fig. 5a).

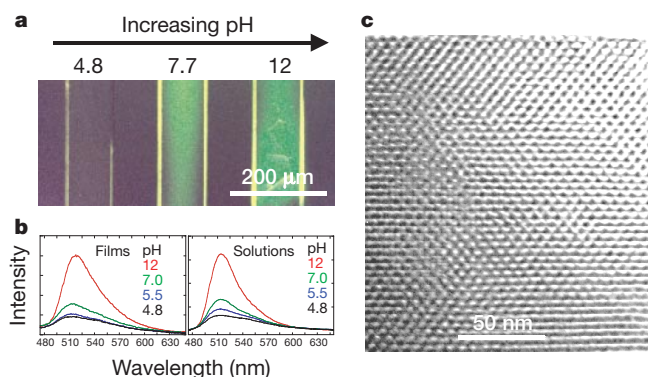


Figure 5 Patterned pH-sensitive fluidic system. **a**, Fluorescence image of three adjacent 5,6-FAM, SE-conjugated pore channel networks after introduction of aqueous solutions prepared at pH 4.8, 7.7 or 12.0. Patterned dye-conjugated thin film mesophases were prepared according to scheme 2 (Fig. 4). Aqueous solutions of varying pH were introduced on the terminal pads (Fig. 4) and transported into the imaging cell by capillary flow. Image was acquired using a Nikon Diaphot 300 inverted microscope and 520-nm bandpass filter. **b**, Fluorescence spectra of 5,6-FAM, SE-conjugated mesoporous films upon exposure to aqueous solutions of pH 4.8, 7.7 and 12.0. Shown for comparison are fluorescence spectra of $0.1 \mu\text{M}$ solutions of 5,6-FAM, SE prepared in aqueous solutions of pH 4.8, 7.7 and 12.0. The similarity of the two sets of spectra confirms the maintenance of dye functionality upon conjugation within the mesoporous channel system. **c**, Cross-sectional TEM micrograph of the patterned, dye-conjugated thin-film mesophase, providing evidence of the 3D pore channel network.

computer-aided design (see Supplementary Information Fig. 4), greater printing resolution achieved compared to standard ink (see Fig. 3a), and our ability to selectively functionalize the ink, suggest applications in sensor arrays and display technologies.

MPL and IJP are serial techniques. In situations where it is desirable to create an entire pattern with the same functionality, it might be preferable to employ a parallel technique in which the deposition process occurred simultaneously in multiple locations. Scheme 2 (Fig. 4) illustrates dip-coating on patterned SAMs. This rapid, parallel procedure uses microcontact printing²⁴ or electrochemical patterning²⁵ of hydroxyl- and methyl-terminated SAMs to define hydrophilic and hydrophobic patterns on the substrate surface. Then, using inks identical to those employed for MPL and IJP, preferential ethanol evaporation during dip-coating enriches the depositing film in water, causing selective de-wetting of the hydrophobic regions and ensuring self-assembly of silica-surfactant mesophases exclusively on the hydrophilic patterns. In this fashion, multiple lines, arrays of dots, or other arbitrary shapes, can be printed in seconds (see Fig. 5 in Supplementary Information). Recent density functional theory calculations²⁶ have established the ultimate resolution of this differential wetting technique to be 1–2 nm.

Scheme 2 (Fig. 4) illustrates the formation of a patterned propyl-amine-derivatized cubic mesophase by selective de-wetting followed by calcination to remove the surfactant templates (the organosilane used was aminopropyltrimethoxysilane—compound 3 in Table 1.) The TEM micrograph (Fig. 4, inset A) and N₂-sorption isotherm (based on surface acoustic wave, SAW, measurements; Fig. 4, inset B) provide evidence of the mesostructural order and proof of the accessibility of the mesoporosity to the vapour phase. In order to make a pH-sensitive fluidic system, the covalently bound propyl-amine ligands were conjugated with a pH-sensitive dye, 5,6-carboxyfluorescein, succinimidyl ester (5,6-FAM, SE) introduced in the pore-channel network of the cubic mesophase. After removal of any non-covalently-bonded dye, the uniform, continuous porosity of the amine-derivatized and dye-conjugated films was confirmed by TEM (Fig. 5c) and the corresponding SAW-based nitrogen sorption isotherm (Fig. 4, inset B). The slight reduction in film porosity after dye conjugation reflects the volume occupied by the attached dye moieties. The patterned, dye-conjugated array was used to monitor the pH of fluids introduced at terminal pads and transported by capillary flow into an imaging cell (scheme 2, Fig. 4).

Figure 5a shows the fluorescence image of an array contacted with three different aqueous solutions prepared at pH 4.8, 7.7 and 12.0. Figure 5b shows the corresponding emission spectra. Comparison with solution data (Fig. 5b) indicates that dye molecules covalently attached to the mesoporous framework retain similar functionality to those in solution. The combined fluorescence image (Fig. 5a) and plan-view and cross-sectional TEM micrographs (Figs 4 and 5c) of the patterned dye-conjugated film demonstrate the uniformity of macro- and meso-scale features achievable by this evaporation-induced, de-wetting and self-assembly route. In comparison, films formed slowly (2–24 h) by nucleation and growth of thin-film mesophases on patterned SAMs² were observed to have non-homogeneous, globular morphologies unsuitable for fluidic or photonic systems. We also note that in this case the mesoporous film formed on the hydrophobic regions.

We believe that the evaporation-induced self-assembly process described here, and its elaboration in three different printing procedures, holds great promise for rapid prototyping of functional fluidic and photonic devices, along with displays and sensor arrays. Compared to alternative approaches like micromoulding in capillaries, printing is considerably faster (seconds compared to 12 hours) and avoids the need for moulds, masks and resists. By using a spectrum of functional inks and interfacing with commercially available software, computer-aided design and rapid transcription of functional microsystems may soon be achievable. □

Methods

Experimental procedures

Precursor solutions used as inks were prepared by addition of surfactants (cationic, CTAB; CH₃(CH₂)₁₅N⁺(CH₃)₃Br[−] or non-ionic, Brij-56; CH₃(CH₂)₁₅-(OCH₂CH₂)₁₀-OH), organosilanes (R'-Si(OR)₃, see Table 1), or organic molecules (see Table 1) to an acidic silica sol prepared from TEOS [Si(OCH₂CH₃)₄] (A2**). The acid concentration employed in the A2** synthesis procedure was chosen to minimize the siloxane condensation rate, thereby promoting facile self-assembly during printing²⁷. In a typical preparation, TEOS, ethanol, water and dilute HCl (mole ratios: 1:3.8:1.5:10^{−5}) were refluxed at 60 °C for 90 min. The sol was diluted with 2 volumes of ethanol followed by further addition of water and HCl. Organosilanes (R'-Si(OR)₃, where R' is a non-hydrolysable organic functional ligand) were added followed by surfactants and (optionally) organic additives (see Table 1). Surfactants were added in requisite amounts to achieve initial surfactant concentrations *c*₀ ranging from 0.004 to 0.23 M (*c*₀ ≪ c.m.c.). The final reactant molar ratios were: 1 TEOS: 22 C₂H₅OH: 5 H₂O: 0.093–0.31 surfactant: 0.039–0.8 organosilanes: 2.6 × 10^{−5} organic additives. For the ethane-bridged silsesquioxane, (RO)₃Si-(CH₂)₂-Si(OR)₃ (6 in Table 1), the neat precursor was diluted in ethanol and mixed with 1–8 wt% CTAB or Brij-56 surfactant followed by addition of an aqueous solution of HCl. The final reactant molar ratios were: Si:EtOH:H₂O:HCl:surfactant = 1:22.5:0.004:0.054–0.18. Co-hydrolysis of organosilanes with TEOS in the initial A2** sol preparation generally resulted in disordered worm-like mesostructures²⁸. After pattern deposition and drying, the surfactant templates were selectively removed by calcination in a nitrogen atmosphere at a temperature sufficient to decompose the surfactant molecules (~350 °C) without degrading the covalently bound organic ligands R' (confirmed by ²⁹Si and ¹³C MAS NMR spectroscopy) or by solvent extraction.

Patterning procedures

Micropen lithography was performed using a Model 400a micropen instrument (Ohmcraft Inc.). The pen orifice was 50 μm and the writing speed was 2.54 cm s^{−1}. The pattern was designed using AutoCAD 14.01 (Autodesk, Inc.) software. Ink-jet printing was performed using a Model HP DeskJet 1200C printer (Hewlett-Packard Co.). The pattern was designed using Microsoft PowerPoint 98 software.

Dip-coating of patterned (hydrophilic/hydrophobic) substrates was performed at a withdrawal speed of 7.6–51 cm min^{−1} under ambient laboratory conditions. Hydrophilic/hydrophobic patterns were created by microcontact printing of hydrophobic, *n*-octadecyltrichlorosilane (CH₃(CH₂)₁₇SiCl₃) SAMs²⁹ on hydrophilic silicon substrates (hydroxylated native oxide) or by a technique involving electrochemical desorption of a hydroxyl-terminated SAM prepared from 11-mercaptoundecanol (HO(CH₂)₁₁SH) from patterned, electrically isolated gold electrodes followed by immersion in a 1 mM ethanolic solution of 1-dodecanethiol, CH₃(CH₂)₁₁SH (ref. 25).

Received 23 August 1999; accepted 6 March 2000.

- Xia, Y. & Whitesides, G. M. Soft lithography. *Angew. Chem. Int. Edn Engl.* **37**, 550–575 (1998).
- Yang, H., Coombs, N. & Ozin, G. A. Mesoporous silica with micrometer-scale designs. *Adv. Mater.* **9**, 811–814 (1997).
- Trau, M. *et al.* Microscopic patterning of orientated mesoscopic silica through guided growth. *Nature* **390**, 674–676 (1997).
- Yang, P. *et al.* Hierarchically ordered oxides. *Science* **282**, 2244–2246 (1998).
- Kresge, C., Leonowicz, M., Roth, W., Vartuli, C. & Beck, J. Ordered mesoporous molecular sieves synthesized by a liquid-crystal template mechanism. *Nature* **359**, 710–712 (1992).
- Burkett, S. L., Sims, S. D. & Mann, S. Synthesis of hybrid inorganic-organic mesoporous silica by co-condensation of siloxane and organosiloxane precursors. *Chem. Commun.* 1367–1368 (1996).
- Fowler, C. E., Burkett, S. L. & Mann, S. Synthesis and characterization of ordered organo-silica-surfactant mesophases with functionalized MCM-41-type architecture. *Chem. Commun.* **18**, 1769–1770 (1997).
- Lim, M. H., Blanford, C. F. & Stein, A. Synthesis and characterization of a reactive vinyl-functionalized MCM-41: Probing the internal pore structure by a bromination reaction. *J. Am. Chem. Soc.* **119**, 4090–4091 (1997).
- Liu, J. *et al.* Hybrid mesoporous materials with functionalized monolayers. *Adv. Mater.* **10**, 161–165 (1998).
- Lebeau, B., Fowler, C. E., Hall, S. R. & Mann, S. Transparent thin films and monoliths prepared from dye-functionalized ordered silica mesostructures. *J. Mater. Chem.* **9**, 2279–2281 (1999).
- Imagaki, S., Guan, S., Fukushima, Y., Ohsumi, T. & Terasaki, O. Novel mesoporous materials with a uniform distribution of organic groups and inorganic oxide in their frameworks. *J. Am. Chem. Soc.* **121**, 9611–9614 (1999).
- Melde, B. J., Holland, B. T., Blanford, C. F. & Stein, A. Mesoporous sieves with unified hybrid inorganic/organic frameworks. *Chem. Mater.* **11**, 3302–3308 (1999).
- Asefa, T., MacLachlan, M. J., Coombs, N. & Ozin, G. A. Periodic mesoporous organosilicas with the channel inside the channel walls. *Nature* **402**, 867–871 (1999).
- Brinker, C. J., Lu, Y., Sellinger, A. & Fan, H. Evaporation-induced self-assembly: nanostructures made easy. *Adv. Mater.* **11**, 579–585 (1999).
- Yang, H., Coombs, N., Sokolov, I. & Ozin, G. A. Free-standing and oriented mesoporous silica films grown at the air–water interface. *Nature* **381**, 589–592 (1996).
- Sellinger, A. *et al.* Continuous, self-assembly of organic–inorganic nanocomposite coatings that mimic nacre. *Nature* **394**, 256–260 (1998).
- Yang, P. *et al.* Direct-write precision resistors for ceramic packages. *Mater. Res. Soc. Symp. Proc.* **542**, 159–164 (1999).
- Piner, R. D., Zhu, J., Xu, F., Hong, S. & Mirkin, C. A. “Dip-pen” nanolithography. *Science* **283**, 661–663 (1999).
- Prakash, S. S., Brinker, C. J., Hurd, A. J. & Rao, S. M. Silica aerogel films prepared at ambient pressure by using surface derivatization to induce reversible drying shrinkage. *Nature* **374**, 439–443 (1995).

20. Yang, P. *et al.* Mirrorless lasing from mesostructured waveguides patterned by soft lithography. *Science* **287**, 465–467 (2000).
21. Chang, S.-C. *et al.* Multicolor organic light-emitting diodes processed by hybrid inkjet printing. *Adv. Mater.* **11**, 734–737 (1999).
22. Pede, D., Serra, G. & De Rossi, D. Microfabrication of conducting polymer devices by ink-jet stereolithography. *Mater. Sci. Eng. C*, **5**, 289–291 (1998).
23. Lu, Y. *et al.* Aerosol-assisted self-assembly of spherical, silica nanoparticles exhibiting hexagonal, cubic and vesicular mesophases. *Nature* **398**, 223–226 (1999).
24. Wilbur, J. L., Kumar, A., Biebuyck, H. A., Kim, E. & Whitesides, G. M. Microcontact printing of self-assembled monolayers: applications in microfabrication. *Nanotechnology* **7**, 452–457 (1996).
25. Tender, L. M., Worley, R. L., Fan, H. & Lopez, G. P. Electrochemical patterning of self-assembled monolayers onto microscopic arrays of gold electrodes fabricated by laser ablation. *Langmuir* **12**, 5515–5518 (1996).
26. Douglas Frink, L. J. & Sallinger, A. G. Wetting of a chemically heterogeneous surface. *J. Chem. Phys.* **110**, 5969–5975 (1999).
27. Lu, Y. *et al.* Continuous formation of supported cubic and hexagonal mesoporous films by sol-gel dip-coating. *Nature* **389**, 364–368 (1997).
28. Tanev, P. T. & Pinnavaia, T. J. A neutral templating route to mesoporous molecular-sieves. *Science* **267**, 865–867 (1995).
29. Bain, C. D. *et al.* Formation of monolayer films by the spontaneous assembly of organic thiols from solution onto gold. *J. Am. Chem. Soc.* **111**, 321–335 (1989).
30. Cairncross, R. A., Schunk, P. R., Baer, T. A., Rao, R. R. & Sackinger, P. A. A finite element method for free-surface flows of incompressible fluids in three dimensions, Part 1: Boundary-fitted mesh motion. *Int. J. Numer. Meth. Fluids* (in the press).
31. Baer, T. A., Cairncross, R. A., Schunk, P. R., Rao, R. R. & Sackinger, P. A. A finite element method for free-surface flows of incompressible fluids in three dimensions, Part 2: Dynamic wetting lines. *Int. J. Numer. Meth. Fluids* (in the press).
32. Frye, G. C., Ricco, A. J., Martin, S. J. & Brinker, C. J. in *Better Ceramics Through Chemistry III* (eds Brinker, C. J., Clark, D. E. & Ulrich, D. R.) 349–354 (Materials Research Society, Reno, Nevada, 1988).

Supplementary Information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank R. Assink for performing NMR experiments, P. Yang for technical assistance with micropen lithography, and K. Burns for technical illustrations. TEM investigations were performed in the Department of Earth and Planetary Sciences at the University of New Mexico. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed-Martin Company, for the US Department of Energy (DOE). This work was supported by the US DOE Basic Energy Sciences Program, the Sandia National Laboratories Laboratory-Directed Research and Development Program, the Defense Advanced Research Projects Agency Bio-Weapons Defense Program, and the US Department of Defense Office of Naval Research.

Correspondence and requests for materials should be addressed to C.J.B. (e-mail: cjbrink@sandia.gov).

Large differences in tropical aerosol forcing at the top of the atmosphere and Earth's surface

S. K. Satheesh & V. Ramanathan

Center for Clouds, Chemistry, and Climate (C⁴), Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California 92093 USA

The effect of radiative forcing by anthropogenic aerosols is one of the largest sources of uncertainty in climate predictions^{1–6}. Direct observations of the forcing are therefore needed, particularly for the poorly understood tropical aerosols. Here we present an observational method for quantifying aerosol forcing to within ± 5 per cent. We use calibrated satellite radiation measurements and five independent surface radiometers to quantify the aerosol forcing simultaneously at the Earth's surface and the top of the atmosphere over the tropical northern Indian Ocean. In winter, this region is covered by anthropogenic aerosols of sulphate, nitrate, organics, soot and fly ash from the south Asian continent^{7,8}. Accordingly, mean clear-sky solar radiative heating

for the winters of 1998 and 1999 decreased at the ocean surface by 12 to 30 W m⁻², but only by 4 to 10 W m⁻² at the top of the atmosphere. This threefold difference (due largely to solar absorption by soot) and the large magnitude of the observed surface forcing both imply that tropical aerosols might slow down the hydrological cycle.

During December to April of each year, the tropical Indian Ocean/Atmosphere system acts as a natural experiment for observing the radiative forcing by anthropogenic aerosols⁷. The gaseous and particulate pollutants emitted by the Indian sub-continent and the south Asian region are transported over the entire north Indian Ocean by the persistent northeastern low-level monsoonal flow reaching as far south as 5° to 10° S (refs 7–11). The Indian Ocean Experiment, INDOEX⁷, was organized to assess the climatic and chemical influence of anthropogenic aerosols in this region. This focused international experiment has been collecting data from ships, satellites and surface stations since 1996, and culminated in the intensive field phase conducted during January to March 1999.

The Kaashidhoo Climate Observatory (KCO)^{7,8}, established as part of INDOEX in the island of Kaashidhoo at 4.965° N, 73.466° E about 500 km southwest of the southern tip of India, has been continually measuring the aerosol chemical, radiative and micro-physical properties since February 1998. During the winter monsoon, the air mass over KCO mostly (about 90% of the time) originates from the Indian and south Asian sub-continent¹¹. Coincident with the KCO observations are the radiation budget data from NASA's Clouds and Earth's Radiant Energy system (CERES) satellite, with an absolute accuracy of 0.2% or better, with a comparable precision^{12,13}. The KCO data, along with several ship-borne measurements, have been documented extensively^{8–11}. The data have also been used to develop a detailed aerosol–radiation model⁸. Here we compare the results of this model with the observed aerosol forcing. The model simulates the observed surface and TOA solar fluxes to within a few per cent (ref. 8).

The aerosol columnar optical depth at 0.5 μ m wavelength (τ_v) ranges from 0.2 to 0.7 and undergoes significant daily, monthly and inter-annual variability (Fig. 1), and is a good index for the overall solar radiative effect of aerosols. The mean value of τ_v shown in Fig. 1 during February–March 1999 was ~ 0.41 compared to the corresponding value of 0.16 during 1998. The variability is much higher at smaller wavelengths ($\sim 0.34 \mu$ m) compared to near-infrared wavelengths ($\sim 1 \mu$ m), indicating that the increase in aerosol optical depth is dominated by sub-micrometre particles typical of anthropogenic sources. The KCO aerosol measurements⁸ reveal that sulphate and ammonium are responsible for 29% of the observed optical depth (τ_v), sea-salt and nitrate about 17%, mineral dust about 15%, and the inferred soot, organics, and fly ash contribute respectively 11%, 20% and 8%. The single scattering albedo, ω_0 (where $\omega_0 = \text{scattering}/[\text{scattering} + \text{absorption}]$), for the composite

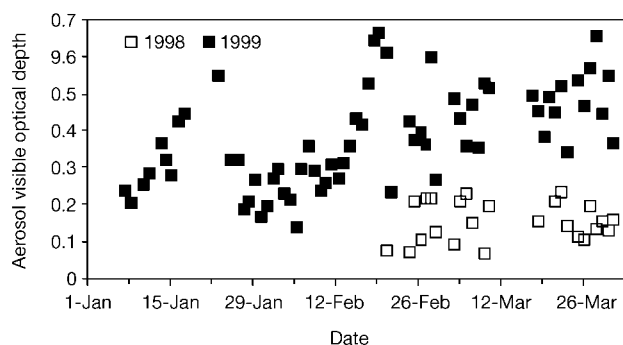


Figure 1 Temporal variation of aerosol optical depth. The figure shows aerosol optical depth at 500 nm wavelength for 1998 (open squares) and 1999 (filled squares). During 1998, the observations are available only during February and March.

aerosol was measured to be consistently in the range of 0.87 to 0.9, regardless of the large fluctuation in optical depths, indicating a highly absorbing aerosol. Anthropogenic sources contribute ~60% or more to the observed optical depths⁸.

We employ two methods, method 1 and method 2, for estimating the surface forcing efficiency, f_e^S , to reduce radiometric uncertainties¹⁴ (see Methods section for radiometers used). In method 1, f_e^S is obtained by subtracting the observed solar flux from that estimated by the model for an aerosol-free atmosphere. The main uncertainty in method 1 is that instrumental offsets or bias in the model will bias the forcing accordingly. In method 2, the aerosol forcing estimated from method 1 is plotted as a function of the observed aerosol optical depth (Fig. 2a). The slope of the forcing versus τ_v yields the aerosol surface forcing efficiency, f_e^S . The value of f_e^S when multiplied by the individual daily value of τ_v yields the aerosol forcing for each day for method 2. The forcing estimated from method 2 is not influenced by instrumental or model offsets.

Several consistency checks assured us of the robustness of the observed f_e^S values of -70 to -75 W m^{-2} . First, these are consistent with modelled values (Fig. 2a), when single-scattering albedos were specified to be within the observed range of 0.87 to 0.9. Second, f_e^S agreed within 5% when obtained separately for the 1998 and 1999

data. The third check comes from the f_e^S obtained from the spectro-radiometer ($0.4\text{--}1.0 \mu\text{m}$), which is about -55 W m^{-2} . When this value is multiplied by the narrow-band to broad-band difference (a factor of ~ 1.3), it is in excellent agreement with the values in Fig. 2a. Another check is with photodiode radiometer ($0.4\text{--}0.7 \mu\text{m}$) measurements, obtained during the 1996 cruise of ORV *Sagar Kanya*, and which yielded⁹ an f_e^S of about -40 W m^{-2} for the northern Indian Ocean. For this value, when multiplied with the modelled ratio of ~ 1.73 for the broadband to visible ($0.4\text{--}0.7 \mu\text{m}$) values of f_e^S , we obtain -69 W m^{-2} . The similarity between 1996, 1998 and 1999 and between KCO and other regions in the northern Indian Ocean indicate that the KCO forcing efficiency of -70 to -75 W m^{-2} is representative of f_e^S for the tropical Indian Ocean when it is subject to polluted air masses.

The four independent observations for the monthly mean forcing and the model values are in excellent agreement (Table 1). The aerosols decrease the ocean solar heating by -23 to -37 W m^{-2} which is equivalent to a reduction by 10–15%. The collocated clear-sky forcing at the top of the atmosphere (TOA), estimated from the CERES radiation budget data^{15,16}, is plotted against aerosol visible optical depths (Fig. 2b) to get TOA forcing efficiency, f_e^T , of about -25 W m^{-2} , which compares well with the model estimate of about -22 W m^{-2} . Our TOA forcing values are also consistent with the Indian Ocean values deduced from the Earth Radiation Budget Experiment data¹⁷. The decrease of $45\text{--}50 \text{ W m}^{-2}$ between TOA and surface forcing f_e^T and f_e^S implies absorption of solar radiation in the atmosphere due to aerosols. The ratio f_e^S/f_e^T , an indication of the aerosol absorbing efficiency, is ~ 3 for the observations (Fig. 2a and 2b) and about 3.4 for the model. A model with just sea-salt or sulphate aerosols will yield a ratio of only 1.5 (refs 17, 18). When soot is removed from the KCO model, the ratio decreases to ~ 1.9 . Thus, although soot contributes only about 10% to τ_v , it strongly affects the surface forcing.

For 1998, the mean forcing values in February and March are: $f_e^S = -12 \text{ W m}^{-2}$ and $f_e^T = -4 \text{ W m}^{-2}$, and the atmospheric forcing, $f^A = f^T - f_e^S$, is $+8 \text{ W m}^{-2}$. The corresponding values for 1999 are: $f_e^S = -30 \text{ W m}^{-2}$, $f_e^T = -10 \text{ W m}^{-2}$ and $f^A = 20 \text{ W m}^{-2}$. How representative are the large forcing values for the rest of the Indian Ocean?

The high optical depth values shown in Fig. 1 are not unique to KCO. Satellite measurements over the tropical Indian Ocean¹⁹ during the 1998 and 1999 winter months reveal that the northern Indian Ocean has τ_v values exceeding 0.2 for latitudes north of 5°N , and τ_v increases northwards reaching values as high as 0.4 to 0.6 towards the coastal oceans. Ship-borne measurements since 1996 corroborate the satellite data^{9,10}. When this is combined with the f_e^S shown in Fig. 2, we infer that the surface solar heating during the winter months may have decreased by as much as $15\text{--}40 \text{ W m}^{-2}$ over the northern Indian Ocean, and the lower atmospheric solar heating increased by $10\text{--}30 \text{ W m}^{-2}$. The atmospheric heating translates into a diurnal mean heating rate perturbation of about $0.3\text{--}1 \text{ K d}^{-1}$ with $1\text{--}3 \text{ K d}^{-1}$ during local noon which is an increase of 50–100% of the aerosol-free solar heating.

The large values of negative surface forcing by aerosols and associated atmospheric heating raise several issues. For absorbing aerosols, the surface forcing and the atmospheric forcing are more

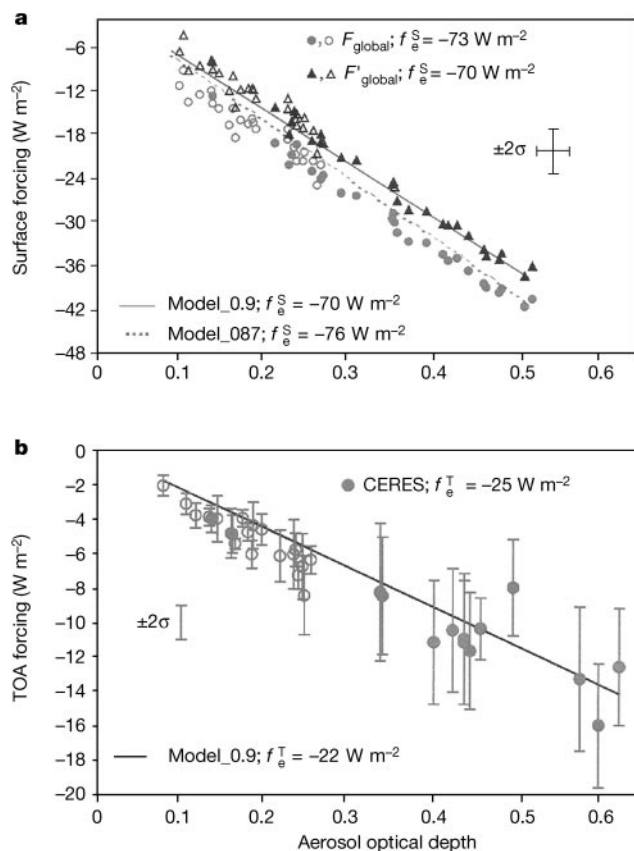


Figure 2 Aerosol forcing in the atmosphere. **a**, The solar aerosol forcing at the Earth's surface (broadband and diurnally averaged) as a function of columnar aerosol optical depth at 500 nm wavelength. The circles correspond to F_{global} (direct + diffuse) and triangles correspond to F'_{global} . Open symbols represent 1998 and solid symbols represent 1999. The vertical and horizontal bars represent the uncertainties in the forcing and optical depths, respectively. The surface forcing efficiency, f_e^S is the slope of the surface forcing with respect to aerosol optical depth. The top of the atmosphere (TOA) forcing efficiency, f_e^T is the slope of the TOA forcing with respect to aerosol optical depth. Model_0.9 and Model_0.87 represent the modelled value of the forcing efficiencies for aerosol single-scattering albedos of 0.9 and 0.87 respectively. **b**, The solar aerosol forcing at the TOA. The vertical bar represents the pixel-to-pixel variability of the TOA fluxes. Open circles represent 1998 and solid circles represent 1999.

Table 1 Comparison of the monthly mean surface aerosol direct forcing

Month	Method	F_{global} flux	F'_{global} flux (W m^{-2})	Model (W m^{-2})
February 1999	Method 1	-28.3	-22.7	-24.1
	Method 2	-24.5	-23.8	
March 1999	Method 1	-37.4	-31.9	-33.3
	Method 2	-34.1	-33.0	

F_{global} is the global flux obtained by adding the direct flux measured by a pyrheliometer and the diffuse flux measured by a shaded pyranometer. F'_{global} is the global flux measured by an unshaded pyranometer. Method 2 is more accurate since it is not influenced by instrumental biases.

relevant climate quantities than the TOA forcing, as also acknowledged elsewhere^{3,20}. Next, how would the tropical oceans respond to such a large decrease in surface solar heating? One possibility is that it might decrease the evaporation (since nearly 80% of the radiative heating in the tropics is balanced by evaporation²¹) thus decreasing the intensity of the hydrological cycle. The ocean may also respond by altering its heat transport. Another major question is whether the low clouds can sustain the intense soot-induced solar heating, or whether they will burn off. The low-level trade cumulus clouds, north of the inter-tropical convergence zone (ITCZ), are embedded within the aerosol layer (Fig. 3), and the low clouds evaporate when subject to enhanced solar heating^{3,22}. Finally, it is uncertain whether the soot-induced heating can modify the monsoonal circulations. This question arises because of the existence of significant north–south gradients in the sooty haze layer across the ITCZ (Fig. 3). Our results raise substantial new questions regarding the regional climate effects.

Lastly, we explore the global significance of the present findings. The large INDOEX surface forcing values are comparable to the -26 W m^{-2} estimated by the Tropospheric Aerosol Radiative Forcing Experiment (TARFOX)⁴ for the western North Atlantic Ocean, off the east coast of the United States. Although TARFOX estimates are based on only a few days of aircraft data during July 1996, the close similarity in the surface forcing between the two oceanic regions both downwind of significant anthropogenic sources is striking. Direct forcing observations are too few to make generalizations, but we can speculate from observed aerosol optical properties. Large aerosol concentrations ($\tau_v > 0.15$) in conjunction with absorbing aerosols ($\omega_0 < 0.95$) are required to obtain the large negative surface forcing and f^S/f^T ratio of 3. Absorbing aerosols with $\omega_0 < 0.95$ are typical of biomass burning^{5,23}, desert dust²⁴, and soot from fossil-fuel burning²⁵. Such emission sources of absorbing aerosols are prevalent in the Northern Hemisphere^{25,26}. With respect to fossil-fuel burning, recent ice-core data in western Europe reveals that black carbon concentrations have tripled since early 1850 (ref. 27). Clearly, we are in the early stages of understanding the effects of these sub-micrometre particles on the temperature and the water cycle of the planet. □

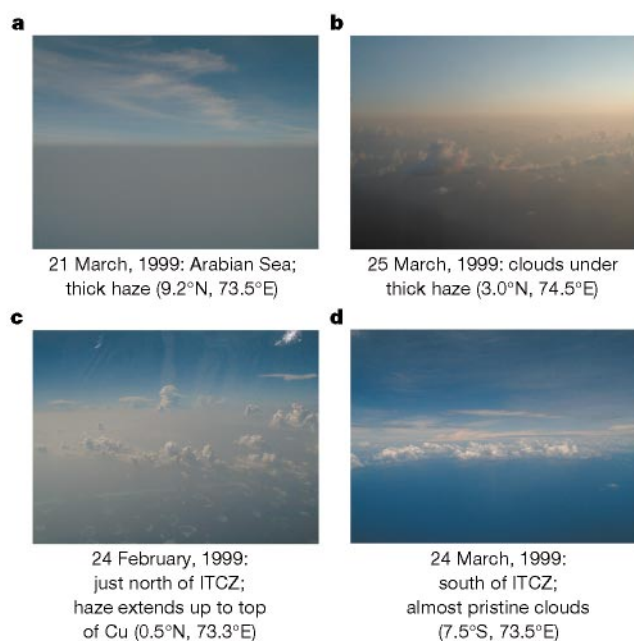


Figure 3 Clouds in the presence of aerosol. Low clouds over the tropical Indian Ocean embedded within the aerosol layer: **a**, **b**, north of the ITCZ; **c**, near the ITCZ; and **d**, free of the haze south of the ITCZ.

Methods

The aerosol optical depths were measured using the CIMEL radiometer of the Aerosol Robotic network (AERONET) of NASA/GSFC as well as a hand-held Microtops radiometer. Because of the inherent uncertainties in the radiometric measurements¹⁴, we used employed two independent broadband instrument sets to obtain the solar irradiance, and two independent analytical techniques to derive the aerosol forcing. In addition, we employed two spectral radiometers to check the broadband instrument results. The direct broadband solar flux (0.2–4 μm) was measured by a Kipp and Zonen pyrliometer mounted on a Sun tracker. The absolute accuracy of this instrument is $\pm 3 \text{ W m}^{-2}$, with a precision of about 1% per year. A ventilated Kipp and Zonen pyranometer, shaded from the direct solar beam, measured the diffuse radiation which, when added to the direct solar flux from the pyrliometer, gave us our best estimate of the global—direct and diffuse—solar flux, denoted by F_{global} . A second independent estimate of the global flux, F'_{global} , was given by an unshaded pyranometer, which measured both the direct and diffuse radiation. The pyranometers have an absolute accuracy of about $\pm 2\%$. In addition, the error in the total pyranometer due to directional response (that is, the detector responds differently to radiation depending on the incident angle) can be as much as $\pm 10 \text{ W m}^{-2}$. A spectro-radiometer (Analytical Spectral Devices) measured the global flux at 512 narrow bands from 0.35 to 1.05 μm (ref. 10). We used the bands from 0.4 to 1.0 μm in the present study. A photodiode radiometer (Biospherical Instruments) measured the global flux from 0.4 to 0.7 μm (ref. 9). For the direct forcing, we need fluxes for clear skies which were obtained using procedures documented elsewhere^{8,10}. The pyranometers and pyrliometer were calibrated at Kipp and Zonen, and also at the Climate Monitoring and Diagnostic Laboratory of the National Oceanic & Atmospheric Administration. The calibration coefficients between the two sources agree within 1%, and they changed by less than 0.5% between 1998 and 1999. The direct flux was corrected for the diffuse flux entering the field of view, and the diffuse flux was corrected for the shading ball and the shading arm. The shading ball shades the Sun to enable measurement of the diffuse flux from the sky. The shading arm, which holds the ball, also blocks the diffuse flux. Together this correction is about 6 W m^{-2} for overhead Sun and 2 W m^{-2} when the solar zenith angle is 60° .

The diurnal average fluxes were obtained by fitting the instantaneous clear-sky flux with a diurnal curve obtained from the aerosol radiation model⁸. The aerosol-free flux was obtained from the radiation model⁸, which employed the observed column water vapour, vertical profiles of ozone, pressure and temperature. Observed ocean surface albedos¹⁵ were employed to convert the downward global flux to net (down minus up) flux at the surface.

The TOA measurements were obtained from CERES on board the Tropical Rainfall Measuring Mission²⁸ satellite which was launched in December 1997. CERES was calibrated pre-launch, and is calibrated routinely within 0.1% using in-orbit flight calibration sources¹². CERES measures the radiances, which are then converted to fluxes using empirical algorithms¹³. The least uncertainty in these algorithms is shown to be for clear-sky oceanic regions¹⁶, which are the only cases considered in this study. The CERES pixel size is 10 km, and CERES fluxes were collocated each day (within $\pm 20 \text{ km}$; 5 min) with KCO optical depth measurement. Surface instruments were used to choose the clear-sky cases. We followed the same procedure outlined under method 1 and method 2 to estimate the diurnal mean forcing at the TOA.

Received 16 October 1999; accepted 31 March 2000.

1. Charlson, R. J. *et al.* Climate forcing by anthropogenic aerosols. *Science* **255**, 423–430 (1992).
2. Houghton, J. T. *et al.* (eds) *Climate Change, 1994: Radiative Forcing of Climate Change and an Evaluation of the IPCC IS92 Emission Scenarios* (Cambridge Univ. Press, Cambridge, 1995).
3. Hansen, J. E. *et al.* Climate forcings in the industrial era. *Proc. Natl Acad. Sci. USA* **95**, 12753–12758 (1998).
4. Russell, P. B. *et al.* Aerosol induced radiative flux changes off the United States mid-Atlantic coast: Comparison of values calculated from sun photometer and in situ data with those measured by air borne pyranometers. *J. Geophys. Res.* **104**, 2289–2308 (1999).
5. Kaufman, Y. J. *et al.* Smoke, Clouds, and Radiation-Brazil (SCAR-B) experiment. *J. Geophys. Res.* **103**, 31783–31808 (1998).
6. Bates, T. S., Huebert, B. J., Gras, J. L., Griffiths, F. B. & Durkee, P. A. International Global Atmospheric (IGAC) Project's first aerosol characterization experiment (ACE 1): Overview. *J. Geophys. Res.* **103**, 16297–16318 (1998).
7. Ramanathan, V. *et al.* Indian Ocean Experiment (INDOEX), A Multi-agency Proposal for Field Experiment in the Indian Ocean (C⁴ publication no. 162, Scripps Institution of Oceanography, La Jolla, 1996); also at (<http://www.indoex.ucsd.edu>) (1996).
8. Satheesh, S. K. *et al.* A model for the natural and anthropogenic aerosols over the tropical Indian Ocean derived from Indian Ocean Experiment data. *J. Geophys. Res.* **104**, 27421–27440 (1999).
9. Jayaraman, A. *et al.* Direct observations of aerosol radiative forcing over the tropical Indian Ocean during the January–February 1996 pre-INDOEX cruise. *J. Geophys. Res.* **D 103**, 13827–13836 (1998).
10. Meywerk, J. & Ramanathan, V. Observations of the spectral clear-sky aerosol forcing over the Tropical Indian Ocean. *J. Geophys. Res.* **104**, 24359–24370 (1999).
11. Krishnamurti, T. N., Jha, B., Prospero, J. M., Jayaraman, A. & Ramanathan, V. Aerosol and pollutant transport and their impact on radiative forcing over tropical Indian Ocean during the January–February, 1996 pre-INDOEX cruise. *Tellus B* **50**, 521–542 (1998).
12. Lee, R. B. *et al.* Pre-launch calibrations of the Clouds and Earth's Radiant Energy System (CERES) tropical rainfall measuring mission and Earth Observing System (EOS): morning (AM-1) spacecraft thermistor bolometer sensors. *IEEE. Trans. Geosci. Remote Sensing* **36**, 1173–1185 (1998).
13. Welicki, B. A. *et al.* Clouds and the Earth radiant Energy System (CERES): An Earth observing system experiment. *Bull. Am. Meteorol. Soc.* **77**, 853–868 (1996).
14. Bush, B. C., Valero, F. P. J., Simpson, A. S. & Bignone, L. Characterization of thermal effects in pyranometers: A data correction algorithm for improved measurement of surface insolation. *J. Atmos. Ocean. Technol.* (in the press).
15. Breigleb, B. P., Minnis, P., Ramanathan, V. & Harrison, E. Comparison of regional clear-sky albedos inferred from satellite observations and model computations. *J. Clim. Appl. Meteorol.* **25**, 214–226 (1986).

16. Ramanathan, V. *et al.* Cloud radiative forcing and climate: Results from the Earth Radiation Budget Experiment. *Science* **243**, 57–63 (1989).
17. Haywood, J. *et al.* Tropospheric aerosol climate forcing in clear-sky satellite observations over the oceans. *Science* **283**, 1299–1303 (1999).
18. Coakley, J. A. & Cess, R. D. Response of the NCAR community climate model to the radiative forcing by the naturally occurring tropospheric aerosol. *J. Atmos. Sci.* **42**, 1677–1692 (1985).
19. Rajeev, K., Ramanathan, V. & Meywerk, J. Regional aerosol distribution and its long range transport over the Indian Ocean. *J. Geophys. Res.* **105**, 2029–2043 (2000).
20. Cess, R. D. *et al.* Climatic effects of large injections of atmospheric smoke and dust—A study of climate feedback mechanisms with one-dimensional and 3-dimensional climate models. *J. Geophys. Res.* **90**, 2937–2950 (1985).
21. Hartman, D. L. *Global Physical Climatology* (Academic, San Diego, 1994).
22. Ackerman, A. S. *et al.* Reduction of tropical cloudiness by soot. *Science* (in the press).
23. Penner, J. E., Dickerson, R. E. & O'Neill, C. A. Effects of aerosol from biomass burning on the global radiation budget. *Science* **256**, 1432–1434 (1992).
24. Tegen, I., Lacis, A. A. & Fung, I. The influence on climate forcing of mineral aerosols from disturbed soils. *Nature* **380**, 419–422 (1996).
25. Cooke, W. F. *et al.* Construction of a 1 degrees × 1 degrees fossil fuel emission data set for carbonaceous aerosol and implementation and radiative impact in the ECHAM4 model. *J. Geophys. Res.* **104**, 22137–22162 (1999).
26. Prospero, J. M. & Nees, R. T. Impact of the north African drought and El-Nino on mineral dust in the Barbados trade winds. *Nature* **320**, 735–738 (1986).
27. Lavanchy, V. M. H., Gaggeler, H. W., Schotterer, U., Schwikowski, M. & Baltensperger, U. Historical record of carbonaceous particle concentrations from a European high-alpine glacier (Colle Gnifetti, Switzerland). *J. Geophys. Res.* **104**, 21227–21236 (1999).
28. Theon, J. S. The Tropical Rainfall Measuring Mission (TRMM). *Adv. Space Res.* **14**, 159–165 (1993).

Acknowledgements

We thank the Government of Maldives for providing the land for KCO and for assistance in its operation. We also thank B. N. Holben for providing the AERONET data, and K. Priestley and N. Loeb for the CERES data. The US component of INDOEX is funded by the NSF and the US Department of Energy, and KCO was funded by the NSF and the Vetlesen Foundation. We also thank J. Fein for providing support to both INDOEX and KCO.

Correspondence and requests for materials should be addressed to V.R.
(e-mail: ram@fiji.ucsd.edu).

Sensitivity of the geomagnetic axial dipole to thermal core–mantle interactions

Jeremy Bloxham

Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138, USA

Since the work of William Gilbert in 1600 (ref. 1), it has been widely believed that the Earth's magnetic field, when suitably time-averaged, is that of a magnetic dipole positioned at the Earth's centre and aligned with the rotational axis. This 'geocentric axial dipole' (GAD) hypothesis has been the central model for the study of the Earth's magnetic field—it underpins almost all interpretations of palaeomagnetic data, whether for studies of palaeomagnetic secular variation, for plate tectonic reconstructions, or for studies of palaeoclimate². Although the GAD hypothesis appears to provide a good description of the Earth's magnetic field over at least the past 100 Myr (ref. 2), it is difficult to test the hypothesis for earlier periods, and there is some evidence that a more complicated model is required for the period before 250 Myr ago³. Kent and Smethurst³ suggested that this additional complexity might be because the inner core would have been smaller at that time. Here I use a numerical geodynamo model and find that reducing the size of the inner core does not significantly change the character of the magnetic field. I also consider an alternative process that could lead to the breakdown of the GAD hypothesis on this timescale, the evolution of heat-flux variations at the core–mantle boundary, induced by mantle convection. I find that a

simple pattern of heat-flux variations at the core–mantle boundary, which is plausible for times before the Mesozoic era, results in a strong octupolar contribution to the field, consistent with previous findings³.

To test the GAD hypothesis directly it is necessary to have palaeomagnetic measurements from known palaeolatitudes (the original latitudes at the time of acquisition of the remanent magnetization which will have changed because of plate tectonics and true polar wander). Then, the palaeomagnetic inclination (I) can be compared with that which would result from a GAD field. Palaeolatitude can be determined for roughly the past 150 Myr from plate reconstructions made using the pattern of magnetic stripes on the sea floor, a process which is largely independent of the validity of the GAD model. Earlier than 150 Myr ago, there is little remaining sea floor, and plate reconstructions become much less complete, so the palaeolatitudes of the samples are poorly determined. Determination of palaeolatitude must then resort to such proxies as palaeoclimate⁴. Although general agreement is found between these proxies and the GAD model², the latitudinal resolution of climate proxies is very limited and departures from the GAD configuration are possible.

An alternative approach to testing the GAD hypothesis is to examine the frequency distribution of $|I|$ (ref. 5). Instead of attempting a test on a site-by-site basis, all the measurements are compiled and a histogram is plotted of frequency versus $|I|$. In a similar vein, here I compare the palaeomagnetic inclination measurements with the cumulative distribution function (c.d.f.) for $|I|$ derived from the GAD model.

In Fig. 1, I compare the $|I|$ c.d.f. for the GAD model with palaeomagnetic measurements from the Global Paleomagnetic Data Base⁶ separated into two time intervals: the Cenozoic and Mesozoic eras (0–250 Myr ago), from which we have selected 3,655 measurements; and the Palaeozoic era and Precambrian time (earlier than 250 Myr ago), from which we have selected 3,503 measurements. We exclude any measurements which are clearly the result of secondary magnetizations. The more recent data conform fairly closely to the GAD model, though with a small deficit of low inclinations. On the other hand, the earlier data show a strong bias towards low inclinations, as previously noted³. A simple measure of this distinction is provided by the median of $|I|$: earlier than 250 Myr ago it is 31°; later than 250 Myr ago it is 48°; and for the GAD model it is 49°.

Several possible explanations have been proposed for the misfit of the older data to the GAD model, including sampling biases, palaeomagnetic artefacts, and departures of the geomagnetic field from the GAD configuration. Kent and Smethurst³ showed that a departure of the field from the GAD configuration in the form of an axial octupole term (of the same sign as the axial dipole and about 25% of its strength) results in a distribution of $|I|$ more closely resembling that from earlier than 250 Myr ago, though they were not able to dismiss other explanations. They proposed that this field configuration may have resulted because the inner core was smaller then than it is at present.

I tested their proposal using a numerical dynamo model. In Fig. 1, I show $|I|$ calculated using a dynamo model⁷ with the Earth's current inner-core radius (approximately 1,200 km) and with a radius 25% smaller. The smaller radius corresponds to a time roughly 1,000 Myr to 2,000 Myr ago depending on the heat flux across the core–mantle boundary⁸. Both the Earth's current inner-core radius and the smaller inner-core radius give distributions which closely match that of the GAD configuration. I conclude from this that the growth of the inner core does not seem to provide an explanation for possible octupolar contributions to the field earlier than 250 Myr ago. Furthermore, my test is conservative in the sense that an inner-core radius has been chosen which represents a time much longer ago than represented by the bulk of the observations from earlier than 250 Myr ago.

It is possible that thermal core–mantle interactions might influence the geodynamo⁹. Mantle convection results in large lateral variations in temperature in the lower mantle which, in turn, result in lateral variations in heat flux across the core–mantle boundary. The pattern of heat flux variations may then affect the pattern of secular variation^{9,10}, and changes in the pattern of heat flux may affect the frequency of geomagnetic reversals¹¹. Support for the former effect comes from models of the historical magnetic field¹² in which persistent concentrations of magnetic flux are observed at high latitude¹³, and are explained with a simple thermal wind model of core–mantle interaction¹⁴. More recent support has come from convection calculations^{15,16} with imposed lateral heat flux variations, and from numerical dynamo models^{17,18}. Support for the latter effect has come from studying the frequency of geomagnetic reversal in a numerical dynamo model with various patterns of heat flux variation¹⁹.

Here I investigated the possibility that changes in the pattern of mantle convection and hence in the pattern of heat flux across the core–mantle boundary, may lead to very long term changes in the morphology of the magnetic field. Lateral heterogeneity in the lower mantle as inferred from seismology is dominated by a Y_2^2 (in the standard nomenclature of spherical harmonics) pattern²⁰. It is not unreasonable to expect that the processes by which the length scale of mantle convection is determined (such as the Rayleigh number) will not have changed significantly during, at least, the Phanerozoic, and thus the pattern will have been dominated by degree 2 harmonics during this period. I then examined the effect of a Y_2^0 pattern of heat flux variation, because I expected that this pattern may be effective in reducing the emergence of poloidal field in equatorial regions (as a result of the thermal wind pattern that it will drive), thus leading to an octupolar contribution to the field.

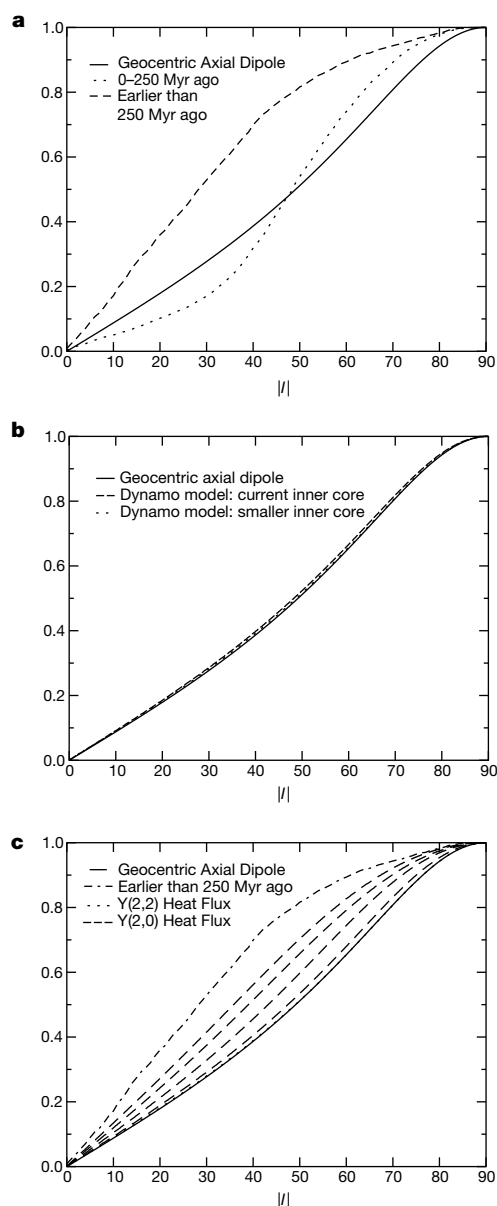


Figure 1 Comparisons of the cumulative distribution function for $|I|$ for the geocentric axial dipole (GAD) with palaeomagnetic measurements and with dynamo models. **a**, Palaeomagnetic measurements taken from the Global Paleomagnetic Data Base⁶ for the intervals 0–250 Myr and earlier than 250 Myr. **b**, The output of a numerical dynamo model with both the present day inner-core radius and with an inner-core radius 75% of the present-day value. **c**, The output of five dynamo model calculations with lateral heat flux variations at the core–mantle boundary. Four of the calculations have a Y_2^0 pattern of

heat flux variation and have r.m.s. amplitude 5%, 11%, 17% and 23% of the superadiabatic heat flux. The smallest-amplitude calculation lies closest to the GAD curve, the largest-amplitude calculation is closest to the palaeomagnetic measurements from earlier than 250 Myr ago. The median inclinations for the dynamo models calculations are 47°, 42°, 39° and 36°, respectively. Also included in this figure is a calculation with a Y_2^2 pattern of heat flux variation, the dominant component in the present-day lower mantle; the curve lies very close to the GAD curve, and is almost entirely obscured by it.

In Fig. 1 I show the results of two sets of calculations. The first, to provide a baseline, is a calculation with a Y_2^0 pattern of heat flux to model the present day mantle. This pattern very closely matches the distribution of $|I|$ for the GAD model (Fig. 1). The second set consists of four calculations each with a Y_2^0 pattern of heat flux but with differing amplitude. These show that the distribution of $|I|$ becomes closer to that observed palaeomagnetically earlier than 250 Myr ago as the amplitude of the variations in heat flux is increased. The calculations all have a non-zero axial octupole term in addition to the axial dipole, with all other terms averaging to zero within one standard deviation. I note that the emergence of an axial octupole term rather than an axial quadrupole could be anticipated given the symmetry properties of the dynamo²¹ (the historical field shows a strong bias towards odd harmonics) and the symmetry of the imposed heat flux variation.

These results demonstrate that by altering the pattern of lateral heat flux variations at the core–mantle boundary it is possible to find dynamos that generate magnetic fields with persistent axial octupoles. In other words, they indicate that the tenability of the GAD model is a function of the pattern of lateral heat flux variations at the core–mantle boundary. Without doubt, that pattern will have been different in the past, so I must conclude that it is unsafe to apply the GAD hypothesis earlier than 250 Myr ago. Indeed, it is no coincidence that the expected timescale for changes in the pattern of mantle convection is similar to the timescale on which we can test the GAD model, as the disappearance of sea floor is itself a manifestation of cycling or turnover of the mantle.

Several cautions must be raised regarding this study. First, because of the uncertain relationship between lowermost mantle heterogeneity and core–mantle boundary heat flux, I am unable to scale the amplitude of the heat flux variations to the Earth, though the amplitudes that I have considered here are not unreasonably large. Second, it is unreasonable to expect that the pattern of heat flux variations would have remained fixed throughout the Precambrian and Palaeozoic. However, suppose that occasionally as the pattern of mantle convection evolves, the lowermost mantle acquires a Y_2^0 signature which gives rise to an axial octupole in the magnetic field, and that at other times (such as the past 250 Myr) the field is close to a geocentric axial dipole. Then, averaging the field over a sufficiently long time interval (one that encompasses periods of Y_2^0 heat flux variation), we would see a bias towards low palaeomagnetic inclination. So, somewhat unexpectedly, the reason that the GAD hypothesis works well for the Cenozoic and Mesozoic may be simply that 250 Myr is too short a period over which to average the field to test the hypothesis adequately. □

Received 9 July 1999; accepted 17 March 2000.

- Gilbert, W. *De Magnete* (London, 1600). Translation: Chiswick Press, London, 1900–1901.
- McElhinny, M. & McFadden, P. *Paleomagnetism* (Academic, San Diego, 2000).
- Kent, D. & Smethurst, M. Shallow bias of paleomagnetic inclinations in the Paleozoic and Precambrian. *Earth Planet. Sci. Lett.* **160**, 391–402 (1998).
- Irving, E. Paleomagnetic and paleoclimatological aspects of polar wandering. *Geofisica Pura e Applicata* **33**, 23–41 (1956).
- Evans, M. Test of the dipolar nature of the geomagnetic field throughout Phanerozoic time. *Nature* **262**, 676–677 (1976).
- McElhinny, M. & Lock, J. IAGA paleomagnetic databases with Access. *Surv. Geophys.* **17**, 557–591 (1996).
- Kuang, W. & Bloxham, J. An Earth-like numerical dynamo model. *Nature* **389**, 371–374 (1997).
- Buffett, B., Huppert, H., Lister, J. & Woods, A. On the thermal evolution of the Earth's core. *J. Geophys. Res.* **101**, 7989–8006 (1996).
- Cox, A. & Doell, R. Long period variations of the geomagnetic field. *Bull. Seismol. Soc. Am.* **54**, 2243–2270 (1964).
- Hide, R. Motions of the Earth's core and mantle, and variations of the main geomagnetic field. *Science* **157**, 3784–3785 (1967).
- Cox, A. The frequency of geomagnetic reversals and the symmetry of the nondipole field. *Rev. Geophys.* **13**, 35–51 (1975).
- Bloxham, J. & Gubbins, D. The secular variation of the Earth's magnetic field. *Nature* **317**, 777–781 (1985).
- Gubbins, D. & Bloxham, J. Morphology of the geomagnetic field and implications for the geodynamo. *Nature* **325**, 509–511 (1987).
- Bloxham, J. & Gubbins, D. Thermal core–mantle interactions. *Nature* **325**, 511–513 (1987).

- Zhang, K. & Gubbins, D. On convection in the Earth's core driven by lateral temperature variations in the lower mantle. *Geophys. J. Int.* **108**, 247–255 (1992).
- Zhang, K. & Gubbins, D. Convection in a rotating spherical fluid shell with an inhomogeneous temperature boundary condition at infinite prandtl number. *J. Fluid Mech.* **250**, 209–232 (1993).
- Sarson, G. R., Jones, C. A. & Longbottom, A. W. The influence of boundary region heterogeneities on the geodynamo. *Phys. Earth Planet. Inter.* **101**, 13–32 (1997).
- Bloxham, J. The effect of thermal core–mantle interactions on the paleomagnetic secular variation. *Phil. Trans. R. Soc. Lond. A* **358**, 1171–1179 (2000).
- Glatzmaier, G., Coe, R., Hongre, L. & Roberts, P. The role of the Earth's mantle in controlling the frequency of geomagnetic reversals. *Nature* **401**, 885–890 (1999).
- Giardini, D., Li, X.-D. & Woodhouse, J. Three-dimensional structure of the Earth from splitting in free-oscillation spectra. *Nature* **325**, 405–411 (1987).
- Gubbins, D. & Zhang, K. Symmetry properties of the dynamo equations for palaeomagnetism and geomagnetism. *Phys. Earth Planet. Inter.* **75**, 225–241 (1993).

Acknowledgements

I thank P. Hoffman, K. Katari, D. Kent, A. Maloof, R. O'Connell, D. Schrag and L. Tauxe for useful conversations during this work. This work was supported by the NSF.

Correspondence and requests for materials should be addressed to the author (e-mail: jeremy_bloxham@harvard.edu).

Early human occupation of the Red Sea coast of Eritrea during the last interglacial

Robert C. Walter*, Richard T. Buffler†, J. Henrich Bruggemann‡§||, Mireille M. M. Guillaume¶, Seife M. Berhe#, Berhane Negassi†*, Yoseph Libsekal**††, Hai Cheng‡‡, R. Lawrence Edwards‡‡, Rudo von Cosel||, Didier Néraudeau§§ & Mario Gagnon|||

* Departamento de Geologica, Centro de Investigación Científica de Educación Superior de Ensenada, km. 107 Carr. Tijuana-Ensenada, Ensenada, B.C., Mexico

† Institute for Geophysics, University of Texas, Austin, Texas 78712, USA

‡ Department of Marine Biology, University of Groningen, PO Box 14, 9750 AA Haren, The Netherlands

§ Department of Marine Biology and Fisheries, University of Asmara, Asmara, PO Box 1220, Eritrea

|| Experimental Zoology Group, Wageningen Institute of Animal Sciences, Wageningen University, PO Box 338, 6700 AH Wageningen, The Netherlands

¶ Muséum National d'Histoire Naturelle, Laboratoire de Biologie des Invertébrés Marins et Malacologie, ESA 8044-CNRS, 55 rue de Buffon, 75005 Paris, France

African Minerals Inc., PO Box 4588, Asmara, Eritrea

†* Department of Mines, Ministry of Energy and Mines, PO Box 272, Asmara, Eritrea

** National Museum of Eritrea, PO Box 5284, Asmara, Eritrea

†† Archaeology Unit, University of Asmara, PO Box 1220, Asmara, Eritrea

‡‡ Department of Geology and Geophysics, University of Minnesota, Minneapolis, Minnesota 55455, USA

§§ Laboratoire de Paléontologie, Géosciences Rennes, Université de Rennes I, Campus de Beaulieu, avenue du Général Leclerc, 35042 Rennes cedex, France

||| Department of Anthropology, University of Toronto, Toronto, Ontario M5S 3G3, Canada

The geographical origin of modern humans is the subject of ongoing scientific debate. The 'multiregional evolution' hypothesis argues that modern humans evolved semi-independently in Europe, Asia and Africa between 100,000 and 40,000 years ago¹, whereas the 'out of Africa' hypothesis contends that modern humans evolved in Africa between 200 and 100 kyr ago, migrating to Eurasia at some later time². Direct palaeontological, archaeological and biological evidence is necessary to resolve this debate. Here we report the discovery of early Middle Stone Age artefacts in an emerged reef terrace on the Red Sea coast of Eritrea, which we date to the last interglacial (about 125 kyr ago) using U–Th

mass spectrometry techniques on fossil corals. The geological setting of these artefacts shows that early humans occupied coastal areas and exploited near-shore marine food resources in East Africa by this time. Together with similar, tentatively dated discoveries from South Africa³ this is the earliest well-dated evidence for human adaptation to a coastal marine environment, heralding an expansion in the range and complexity of human behaviour from one end of Africa to the other. This new, widespread adaptive strategy may, in part, signal the onset of modern human behaviour, which supports an African origin for modern humans by 125 kyr ago.

Our discovery of early Middle Stone Age (MSA) tools within an emerged reef terrace on the Red Sea coast of Eritrea (Fig. 1) expands on earlier reports of artefacts found on the surface of Pleistocene marine deposits elsewhere on the African coast of the Red Sea^{4–6}. Africa is the main source of fossils and artefacts bearing on the study of early hominid evolution, but the Middle to Upper Pleistocene—a period which has great potential for resolving the debate about the origins of modern humans—is largely unstudied as a result of the paucity of terrestrial exposures of the right age⁷. Now geological studies of near-shore marine deposits on the Red Sea coast of East Africa are helping to fill this gap⁸ by showing that such coastal environments can yield important information about early human origins. Our discovery of bifacial handaxes—large, flat, teardrop-shaped tools flaked over at least part of both surfaces—together

with obsidian flake and blade tools is unusual, indicating that old and new tool technologies were used at the time this reef formed.

The area we studied lies along the west coast of the Buri Peninsula (Fig. 1), which exposes a Pleistocene reef terrace near the village of Abdur. This terrace encompasses a 6.5 km by 1 km stretch of raised shoreline on the eastern coast of the Gulf of Zula (Fig. 2a). It is the fossil remnant of a N–S trending fringing reef, which consists of near-shore environments to the east, a central belt of fossil coral reefs, and offshore environments to the west. We subdivided Abdur into three geographic districts: Abdur South (AS), Abdur Central (AC) and Abdur North (AN) (Fig. 2a).

This Pleistocene reef terrace, termed here the Abdur Reef Limestone (ARL), originally formed in shallow water as a shell and coral biostrome, a laterally extensive sheet-like mass of rock built by sedentary marine organisms and composed mainly of their remains. The top of the terrace now ranges in average elevation from about +6 m (above sea level) at Abdur South to +10 m at Abdur Central to +14 m at Abdur North. This general rise in elevation is due, in part, to regional NNW–SSE trending normal faults that step up, en echelon, to the northeast (Fig. 2a). A second group of smaller normal faults is observed that trend NNE–SSW, which might be related to local doming of the Abdur Volcanic Complex that underlies the ARL to the east. This doming apparently also caused seaward tilting of the terrace; about 5° at Abdur South and Abdur Central, and about 1–2° at Abdur North.

The ARL is composed of 1–5 m thick build-ups of molluscs, corals, echinoderms and bioclastic sands. At Abdur North the reef is built mainly on older sedimentary rocks, named here the Buri Sequence, which consist of a succession of shallow marine, fluvial and lacustrine sediments more than 10 m thick that were faulted and deformed prior to deposition of the ARL (Fig. 2a and b). To the east, along its landward fringes, the reef laps onto the Abdur Volcanic Complex (Fig. 2a), which in turn overlies the Buri Sequence. To the south, at Abdur Central and Abdur South, the reef laps onto the volcanic complex, but the underlying Buri Sequence is not exposed.

Abdur North is the most laterally extensive and best studied of the

Table 1 U–Th dated last interglacial terraces from the Red Sea coast of the Sinai Peninsula and Africa

Reference	Location		Terrace height (m asl)		U–Th age (kyr)	
	Name	Latitude	min	max	min	max
11	Sinai					
	Suweira	28° 17'	5	18	119	134
	Mureikha	27° 56'	13	18	125	132
	Point 28	27° 53'	5	8	120	
	Umm-Sid	27° 50'	13	18	81	127
12	Gulf of Aqaba	27° 43'	3	6	118	125
6	Egypt					
	Ras Dib	28° 05'	13	18	115	127
	Gebel Zeit	27° 55'	12	18	117	131
	Zeituna	27° 50'	11		119	126
	Ras Bahar	27° 48'	5	8	123	126
	Ras Gemsa	27° 40'	10	14	125	
	Q. el Qadim	26° 10'	7	8	121	
	S. el Bahari	25° 55'	7		121	
	Wadi Nahari	25° 00'	6		118	
	Wadi Igla	25° 15'	10		115	
	W. K. el Bahari	24° 40'	8		120	
13	Red Sea Islands					
	Brothers	26° 18'	6	8	125	
	Zabargad	23° 37'	6	8	126	
14	Sudan					
	Marub	?	3		134	
	Arkai	20° 10'	3		128	
	Awair	20° 08'	4		132	
	Arus	20° 00'	3		142	
	Irayes	19° 55'	2	6	125	136
15	Eritrea					
	Dahlak Kebir	15° 45'	5		117	170
This study	Abdur	15° 09'	6	14	118	136
16	Djibouti					
	Fagal	12° 26'	3		110	125
	Wayed Kibo	11° 57'	3	12	124	133
	Helento	11° 52'	13		123	
17	Baie Beheta	?	1	9	105	120
	Kala 'assa	13° 10'	2	3	137	
	Tadjura	11° 50'	32		116	
	Obock	11° 56'	20		136	
	Djibouti	11° 30'	9		126	142

m asl, metres above sea level. These dates indicate modest, but variable, tectonic uplift of the Red Sea coast over the last 125 kyr, accounting for current terrace elevations that are slightly above the estimated height of sea level during the last interglacial period: estimates for the height of the Red Sea during the last interglacial vary from +3 m to +8 m (refs 6, 14 and 18).

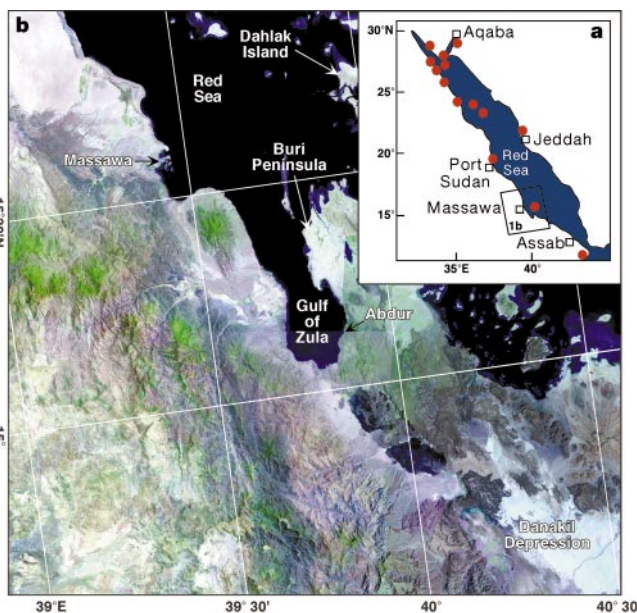


Figure 1 The study area. **a**, Map of the Red Sea basin after ref. 18. Filled circles depict approximate locations of last interglacial coral terraces of the Red Sea (see Table 1). **b**, A 7-4-2 Landsat Thematic Mapper image of the study area. The Abdur Reef Limestone and the archaeological sites discussed here are located on the Gulf of Zula coast near the village of Abdur (arrow). (Scale: 30 minutes in latitude = 55 km). Figure courtesy of M. Abdelsalam, University of Texas at Dallas.

three geographic districts (Fig. 2a). Thirteen stratigraphic sections were measured, and faunal samples were collected from seven representative sections. Here the ARL is composed of the 0.5–3.1 m thick 'lower shell zone', which grades upward into the 0.5–1.5 m thick 'upper coral zone'. Present over much of the area at the base of the ARL, immediately below the lower shell zone, is a distinctive 15–30 cm thick cobble lag deposit consisting of locally derived limestone and volcanic clasts in a carbonate sand matrix. The 'basal cobble zone' overlies an irregular erosional surface, an angular unconformity, that truncates the underlying, tilted Buri Sequence. This lag formed during marine transgression when high-energy wave activity winnowed away finer particles, leaving the coarse cobbles behind. As water depth increased during the transgressive cycle the cobbles submerged into a restricted, quiet bay as an anchorable substrate suitable for growth of rich oyster beds.

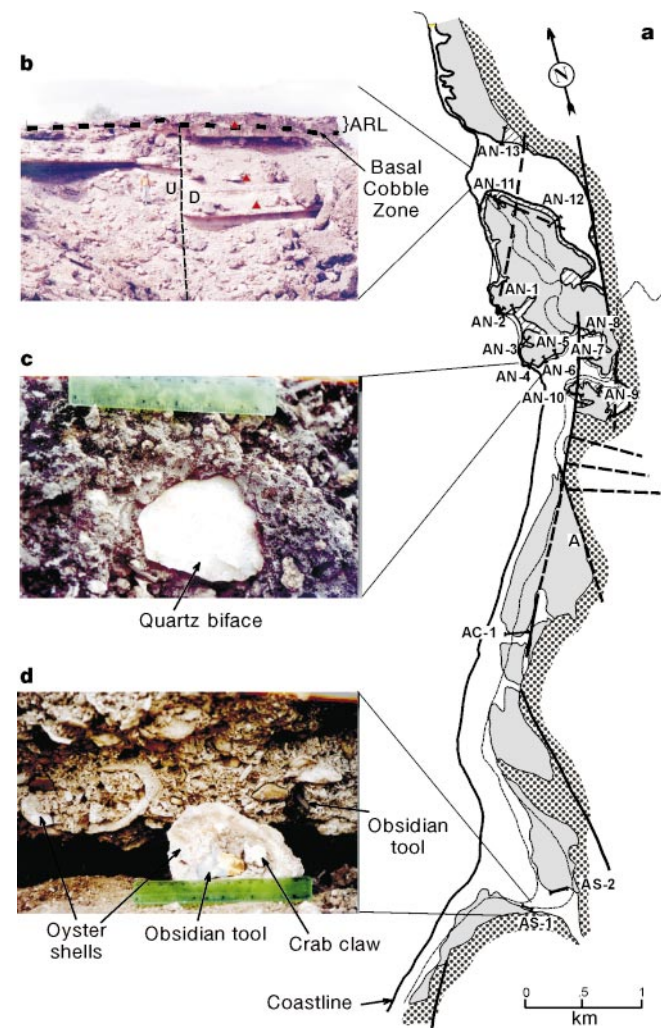


Figure 2 Archaeological finds and their locations. **a**, Geological map of the Abdur field area, showing the location of measured stratigraphic sections (for example, AN-1). Abdur Reef Limestone (ARL), shaded; Abdur Volcanic Complex, dots; Buri Sequence, crosshatch; Faults, heavy lines (dashed where inferred); A, location of Abdur village. **b**, View of section AN-11, showing the stratigraphic position of *in situ* obsidian MSA tools within the basal cobble zone and in the underlying Buri Sequence (triangles). Horizontal dotted line indicates the basal cobble zone. Vertical dashed line indicates a small, normal fault shown along strike (U, up; D, down); the prominent resistant layers are displaced by roughly 3 m across the fault. **c**, A quartz biface, found *in situ* in the lower shell zone of the ARL at AN-4. Scale is 15 cm. **d**, Obsidian tools (one heavily patinated), with oyster and crab parts, all found *in situ* in a coarse bioclastic sand at the base of the lower shell zone at AS-1. Scale is 15 cm.

The lower shell zone is composed of bivalves, gastropods, echinoids, and occasional crustaceans and corals. This zone varies from being nearly all whole shells cemented together without matrix, to being entirely composed of sand- and granule-sized fragments of broken shells. This variation occurs from location to location and vertically within sections. The ecology of mollusc and echinoid species suggests a soft, sandy substrate in a shallow to medium subtidal environment. Bivalves (*Tridacna maxima* and *Spondylus cf. marisrubris*) and gastropods (*Chicoreus ramosus*, *Cypraea limacina* and *Cypraea pantherina*) from the upper part of the lower shell zone indicate proximity to coral reefs.

The upper coral zone is composed of scleractinian corals together with subordinate bivalves, gastropods and echinoids. We interpret this zone to be a fringing reef that formed in a calm embayment, similar to reefs forming today in Massawa Bay (offshore of the port city of Massawa, Fig. 1b). The corals of this zone are: (1) scattered corals not in growth position (for example, Faviidae in AN-4), probably representing a sandy talus seaward of the fringing reef, in water depths of 3 to >8 m; or (2) fields of branching corals in growth position (for example, *Stylophora* sp., *Goniopora* sp., *Galaxea* spp. in AN-7), typical of a reef flat zone of a fringing reef system, in water depths of 0.5–3 m. This linear reef flat extends from AN-13 to AN-10 (Fig. 2a), in a N–S belt, parallel to the palaeoshoreline. It reappears to the south at AC-1 where it shows its best development.

Along the eastern edge of the ARL, near the ancient shoreline of the Buri Peninsula, the shell and coral zones give way laterally to a beach facies, which is 1–2 m thick, resistant carbonate-cemented coarse sand and gravel composed of broken shell fragments interspersed with locally derived gravels of volcanic rock. Scattered shells of oysters, giant clams and crustaceans are found in this facies, as are fossils of terrestrial mammals such as elephant, hippopotamus, rhinoceros and bovid (species not determined).

Abundant stone tools are found *in situ* in the basal cobble zone, the lower part of the lower shell zone and the beach facies of the ARL. At Abdur South, flakes and blades are found throughout the lower shell zone of section AS-1 (Fig. 2d), which here is a bioclastic sand with rich concentrations of separated valves of oysters and other bivalves, gastropods and crustaceans. At AS-2 and AC-1, *in situ* flakes and blades are found in the beach facies. At Abdur North *in situ* bifaces (similar in appearance to Acheulian hand axes^{7,9,10}) are found at AN-1, AN-4 and AN-12, and flakes and blades are found *in situ* in the basal cobble zone, the lower part of the lower shell zone and/or in the beach facies at AN-1, AN-4, AN-7, AN-10, AN-11, AN-12 and AN-13. Bifaces are made primarily from fine-grained volcanic rocks, but also from chert and quartz (Fig. 2c). Flake and blade tools are made mainly from obsidian (Fig. 2d), but occasionally from chert and quartz. Potential sources of these raw materials are found within 1 to 20 km of Abdur.

Our initial estimate of the age of the tool-bearing ARL was 125 kyr, based on its relative height above present sea level. Coral terraces from the Red Sea coast of southern Sinai¹¹, Egypt^{6,12}, the Red Sea Islands¹³, Sudan¹⁴, the Dahlak Islands¹⁵ and Djibouti^{16,17} (Fig. 1a), all of which are about the same elevation as the ARL, date to around 125 kyr by U–Th alpha counting techniques (Table 1). The Red Sea is one of the best-dated Quaternary basins in the world. Fossil beaches and coral reefs of this age (about 115–135 kyr) were the most common landforms on the Red Sea coast¹⁸ when sea level was 3–8 m higher than today (Table 1).

To confirm this estimate, approximately 100 coral samples were collected from the upper coral zone of the ARL for species identification and evaluation for dating. Twenty-four coral samples were subsequently selected for X-ray diffraction studies as possible candidates for U–Th mass spectrometry analyses. Six of the 24 samples yielded aragonite contents between 95% and 100%, all from Abdur North, suggesting that these corals could yield reliable U–Th ages. Four of the samples are from AN-4, one from AN-7 and

one from AN-13 (Fig. 2a), representing a wide distribution across the Abdur North terrace. These six samples were dated using thermal ionization mass spectrometry (TIMS) U–Th techniques (Table 2, refs 19 and 20).

Four coral samples from section AN-4 have primary uranium concentrations and initial uranium isotopic compositions similar to the modern marine value²⁰, suggesting that the dates are accurate. The dates range from 118 to 126 kyr ago, confirming that the ARL formed during the last interglacial high-stand^{19,21,22}, oxygen isotope substage 5e. A fresh quartz hand axe occurs *in situ* in the lower part of the lower shell zone of AN-4 (Fig. 2c), as well as several obsidian flake tools. Obsidian flake tools were also found at AN-4 in the basal cobble zone immediately below the lower shell zone.

The coral sample from section AN-7 yields a date of 136 kyr, slightly older than the dates from AN-4, but still within the temporal range for the last interglacial high-stand compared to other terraces on the Red Sea coast (Table 1). Obsidian flake tools were found in the lower part of the lower shell zone of section AN-7, stratigraphically below the dated coral.

The sample from section AN-13 yields a date of 156 kyr. This is probably erroneous because the uranium concentration is about half the lowest value for primary coralline aragonite (Table 2), which indicates either U loss from the coral or that the U–Th measurement was made on a portion of the coral that had a higher calcite (lower U) abundance than the portion analysed by X-ray diffraction. Either scenario invalidates the date. An erroneous age is suggested because the date of 156 kyr would place the coral in the period of the penultimate glaciation (oxygen isotope stage 6), when sea levels were about 130 m lower than today²³.

Five of six U–Th dates measured on *in situ* corals from Abdur North indicate that the ARL was deposited between 118 and 136 kyr (Table 2). This is consistent with U–Th ages measured on corals from the last interglacial highstand from the Red Sea basin (by alpha counting, Table 1) and from elsewhere (by TIMS^{19,21,22}). A simple mean age for the Abdur Reef Limestone, calculated from the five best U–Th results, is 125 ± 7 kyr (1σ).

We conclude that the stone tools within the ARL are 125 ± 7 kyr old, as shown by their stratigraphic context within this last interglacial reef terrace. Even earlier coastal occupation at Abdur is suggested by the discovery of obsidian tools in the Buri Sequence below the ARL (Fig. 2b), but the age of these deposits is not yet known. The tools in the ARL have fresh, sharp edges, and they are found in low-energy environments. This indicates that the tools were not washed in from older sediments. Neither are they younger artefacts that ‘filtered down’ into the terrace after subaerial exposure, since the tools are found in the lower shell and cobble zones of the terrace and not in the upper coral zone. Such filtering is also precluded by the closely packed nature of the deposit itself.

The primary context is confirmed by the distribution of fresh stone tools within and throughout several reef facies over a wide area (Fig. 2a). Fresh bifaces and obsidian flakes and blades are found

within the lower half metre of the lower shell layer (Figs 2c and d). The tools in the basal cobble zone have fresh, sharp edges, whereas the cobbles are subrounded. Hundreds of freshly preserved obsidian flake and blade tools are found in the beach rock, usually in close proximity to fossil remains of land mammals or near the massive oyster beds.

The simplest explanation for the occurrence of stone tools in the ARL is that whoever made the tools used them to harvest edible (energetically profitable and palatable) shallow marine molluscs and crustaceans and to butcher large land mammals near the shore of this reef system, and then discarded the tools at the various food processing sites. Two edible species of oysters are found in growth position as beds in the basal cobble and lower shell zones (*Hyotissa hyotis* and *Ostrea cf. deformis*). Thirty-one species of edible molluscs were identified from the lower shell zone. Two groups of decapod crustaceans were found in the basal cobble zone and the lower shell zone; one of these groups comprises big brachyuran crabs, which are edible.

Flake and blade tools from the ARL fit into the broad range of variability that defines the early Middle Stone Age (MSA)⁷, but their stratigraphic association with Acheulian-type bifaces is unusual. Early MSA tools in Africa are generally found at sites where such bifaces are no longer present. However, there is currently no precise definition of what tools comprise the earliest MSA⁷ tool-kit, or whether the biface—a tool that is generally associated with much older sites—is or is not part of this transitional assemblage.

The discovery of Palaeolithic artefacts in the Abdur Reef Limestone confirms and expands earlier reports of Acheulian-type tools found near emerged reef terraces from the central Danakil Depression⁴ and the Egyptian coast of the Red Sea^{5,6}. These earlier discoveries were not found in geological context, and consequently no firm assessment of their age is possible. The U–Th dates for the ARL artefacts, however, now permit us to link the Red Sea coast of Eritrea with several South African MSA sites that share evidence of stone tools associated with marine resource adaptations: Herold’s Bay, Sea Harvest, Hoedjies Punt, Klasies River Mouth and Die Kelders’ Cave^{9,24,25}. The oldest of these, Klasies River Mouth and Herold’s Bay, are tentatively dated to 115–100 kyr or correlated with oxygen isotope stage 5 based on geological grounds^{3,24}. Archaic *Homo sapiens* were recently found at a coastal site in Greece, estimated to be about 200 kyr, but no stone tools are linked with the hominids and the date is speculative²⁶.

Archaeological evidence from the ARL confirms that the Red Sea coast of Africa was occupied by early humans by at least 125 kyr ago, but we cannot say, as yet, to which hominid species the tools belong. Human fossils from the last interglacial period at Klasies River Mouth are attributed to near or fully anatomically modern *Homo sapiens*²⁴. Fossils from terrestrial sites of the East African interior, such as Lake Eyasi (~130 kyr), Mumba Rockshelter (~109–130 kyr), the Ngaloba Beds (~120 kyr) and Omo-Kibish (~130 kyr)⁷, suggest that late archaic *Homo sapiens* and/or anatomically modern

Table 2 ²³⁰Th ages and related data for aragonite coral samples from the Abdur Reef Limestone

Sample	Aragonite (%)	²³⁸ U (p.p.b.)	²³² Th (p.p.t.)	²³⁰ Th/ ²³⁸ U (activity)	²³⁰ Th Age (kyr)	$\delta^{234}\text{U}$ (initial)
AN-4-1a no. 1	99	3167 (4)	1309 (11)	0.7550 (19)	118.9 (0.6)	167.5 (1.8)
AN-4-1a no. 1	99	3287 (4)	374 (10)	0.7492 (20)	117.2 (0.6)	167.3 (1.8)
				Average 118.1 (1.0)		
AN-4-1a no. 2	100	3071 (4)	101 (7)	0.7678 (19)	124.0 (0.6)	160.7 (1.6)
AN-4-13a no. 1	96	3153 (4)	991 (12)	0.7537 (18)	119.0 (0.6)	164.8 (1.7)
AN-4-13a no. 3	98	3433 (5)	46 (7)	0.7768 (20)	125.9 (0.7)	166.1 (1.8)
AN-7-6a no. 2	98	2400 (3)	437 (9)	0.8111 (20)	136.4 (0.7)	172.1 (2.0)
AN-13-16a no. 6	95	1280 (1)	4025 (15)	0.8597 (20)	156.0 (0.9)	170.0 (1.9)

Half lives are those quoted in ref. 20. Numbers in parentheses are 2σ errors. A simple mean age for the Abdur Reef Limestone, calculated from the five best U–Th results, is 125 ± 7 kyr; see text for details. Inclusion of the 156 kyr date, which we consider to be erroneous, would make the spread of ages larger than expected for oxygen isotope substage 5e compared to other terraces from the Red Sea (Table 1), except for the Dahlak Islands—the closest of the previously dated terraces to Abdur (Fig. 1 and Table 1).

humans were in East Africa by the time the ARL was forming. Thus, the stone tools from the ARL overlap in time with the apparent transition from archaic to modern *Homo sapiens* in Africa.

Regardless of which hominid species made the tools at Abdur, or at the more tenuously dated sites in South Africa, the pressing question is: what caused such an apparently sudden and widespread coastal marine adaptation by early humans by the last interglacial period? Climate changes leading to hyper-arid conditions in Africa and the Red Sea basin during the penultimate glaciation (~150 kyr) and at the peak of the last interglacial^{6,27,28} may have been severe enough to pressure early humans to migrate from once stable interior habitats to exploit coastal marine habitats for survival. As opposed to shrinking freshwater environments during these times (for example, East African rivers and lakes), human adaptation to marine shoreline environments might have been further influenced by the lack of competition from other terrestrial mammals for coastal marine food resources. It is likely, however, that during extreme hyper-arid conditions marking the peak of the penultimate glaciation, that the Red Sea basin itself was virtually uninhabitable.

Our working model is that: (1) an adaptation to coastal marine environments in Africa by 125 kyr marks a significant, new behaviour for early humans; (2) early humans adopted this strategy in response to environmental stresses caused by fluctuating glacial-interglacial climate cycles that were especially pronounced during the late Middle and early Upper Pleistocene; (3) the proliferation of this strategy by the last interglacial suggests that its first appearance was even earlier; (4) the eventual dispersal of humans out of Africa was due to increased human competition for marine resources, possibly during hyper-arid conditions; and (5) human populations dispersed along the Red Sea coast out of Africa into the Levant by 100 kyr ago²⁹, and perhaps across a landbridge between Africa and Arabia³⁰ at the southern end of the Red Sea that was exposed during sea-level lowstands that mark the last two glacial maxima²³. The artefacts from the Abdur Reef Limestone are in the right geographical location and of the right age to suggest that the route out of Africa was along the coast of Eritrea. It is now clear that the Red Sea basin, and perhaps the entire eastern shoreline of Africa, must be examined for additional evidence to test this model of early human occupation of coastal marine environments before 100 kyr ago, and its role in the dispersal of early humans out of Africa. □

Received 13 December 1999; accepted 17 February 2000.

1. Thorne, A. & Wolpoff, M. The multiregional evolution of humans. *Sci. Am.* April, 28–33 (1992).
2. Stringer, C. B. The emergence of modern humans. *Sci. Am.* December, 98–104 (1990).
3. Deacon, H. J. in *The Human Revolution: Behavioural and biological perspectives on the origins of modern humans* (eds Mellars, P. & Stringer, C.) 547–564 (Princeton Univ. Press, 1989).
4. Faure, H. & Roubet, C. Découverte d'un biface Acheuléen dans les calcaires marins du golfe Pléistocène de l'Adar (Mer Rouge, Éthiopie). *C. R. Acad. Sci. Paris* **267**, 18–21 (1968).
5. Montenat, C. Un aperçu des industries préhistoriques du golfe de Suez et du littoral Égyptien de la Mer Rouge. *Bull. Inst. Fran. d'Archéol. Orient.* **86**, 239–255 (1986).
6. Plaziat, J.-C. et al. in *Sedimentation and Tectonics in Rift Basins: Red Sea-Gulf of Aden* (eds Purser, B. H. & Bosence, D. W. J.) 537–573 (Chapman & Hall, London, 1998).
7. Clark, J. D. The Middle Stone Age of East Africa and the beginnings of regional identity. *J. World Prehist.* **2**, 237–305 (1988).
8. Walter, R. C., Buffler, R. T., Berhe, S., Vondra, C. & Yemane, T. in *Lithospheric Structure, Evolution And Sedimentation In Continental Rifts (Abstracts)* (eds Jacob, A. W. B., Delvaux, D. & Khan, M. A.) (Dublin Institute for Advanced Studies, Dublin, Ireland, 1997).
9. Phillipson, D. W. *African Archaeology* (Cambridge Univ. Press, 1993).
10. Van Peer, P. The Nile corridor and the out-of-Africa model. *Curr. Anthropol.* **39**, 115–140 (1998).
11. Gvirtzman, G., Kronfeld, J. & Buchbinder, B. Dated coral reefs of the southern Sinai (Red Sea) and their implication to late Quaternary sea levels. *Mar. Geol.* **108**, 29–37 (1992).
12. El-Asmar, H. M. Quaternary isotope stratigraphy and paleoclimate of coral reef terraces, Gulf of Aqaba, south Sinai, Egypt. *Quat. Sci. Rev.* **16**, 911–924 (1997).
13. Hoang, C. T. & Taviani, M. Stratigraphic and tectonic implications of uranium-series dated coral reefs from uplifted Red Sea islands. *Quat. Res.* **35**, 264–273 (1991).
14. Hoang, C. T., Dalongeville, R. & Sanlaville, P. Stratigraphy, tectonics and paleoclimatic implications of uranium-series dated coral reefs from the Sudanese coast of the Red Sea. *Quat. Int.* **31**, 47–51 (1996).
15. Conforto, L., Delitala, M. C. & Taddeucci, A. Datazioni col ²³⁰Th di alcune formazioni coralligene delle Isole Dahlak (Mar Rosso). *Soc. It. Min. Pet.* **32**, 153–158 (1976).
16. Hoang, C. T., Lalou, C. & Faure, H. Les récifs soulevés à l'ouest du golfe d'Aden (T. F. A. I.) et les hauts niveaux de coraux de la dépression de l'Adar (Éthiopie), géochronologie et paléoclimats interglaciaires. *Col. Int. CNRS* **219**, 103–116 (1974).
17. Faure, H., Hoang, C. T. & Lalou, C. Datations ²³⁰Th/²³⁴U des calcaires coralliens et mouvements verticaux à Djibouti. *Bull. Soc. Géol. Fr.* **22**, 959–962 (1980).

18. Taviani, M. in *Sedimentation and Tectonics in Rift Basins: Red Sea-Gulf of Aden* (eds Purser, B. H. & Bosence, D. W. J.) 574–582 (Chapman & Hall, London, 1998).
19. Edwards, R. L., Chen, J. H. & Wasserburg, G. J. ²³⁸U-²³⁴U-²³²Th systematics and the precise measurement of time over the past 500,000 years. *Earth Planet. Sci. Lett.* **81**, 175–192 (1986/87).
20. Cheng, H. et al. *The half-lives of uranium-234 and thorium-230* (in the press).
21. Chen, J. H., Curran, H. A., White, B. & Wasserburg, G. J. Precise chronology of the last interglacial period: U-234-Th-230 data from fossil coral reefs in the Bahamas. *Geol. Soc. Am. Bull.* **103**, 82–97 (1991).
22. Stirling, C. H., Esat, T. M., Lambeck, K. & McCulloch, M. T. Timing and duration of the Last Interglacial: evidence for a restricted interval of widespread coral reef growth. *Earth Planet. Sci. Lett.* **160**, 745–762 (1998).
23. Gvirtzman, G., Buchbinder, B., Sneh, A., Nir, Y. & Friedman, G. Morphology of the Red Sea fringing reefs: a result of the erosional pattern of the last-glacial low-stand sea level and the following Holocene recolonization. *Mém. Bur. Rech. Géol. Min.* **89**, 480–491 (1977).
24. Klein, R. G. in *The Human Revolution: Behavioural And Biological Perspectives On The Origins Of Modern Humans* (eds Mellars, P. & Stringer, C.) 529–546 (Princeton Univ. Press, New Jersey, 1989).
25. Brink, J. & Deacon, H. A study of a last interglacial shell midden and bone accumulation at Herold's Bay, Cape Province, South Africa. *Paleoecol. Afr.* **15**, 31–39 (1982).
26. Leney, M. The road out of Africa. *Discov. Archaeol.* **1**, 18 (1999).
27. Freyter, R., Baltzer, F. & Conchon, O. A Quaternary piedmont on an active margin: the Egyptian coast of the NW Red Sea. *Z. Geomorph. N. F.* **37**, 215–236 (1993).
28. Horowitz, A. Continuous pollen diagrams for the last 3.5 m.y. from Israel: vegetation, climate and correlation with the oxygen isotope record. *Paleogeogr. Paleoclimatol. Paleoecol.* **72**, 63–78 (1989).
29. Grün, R. & Stringer, C. B. Electron spin resonance dating and the evolution of modern humans. *Archaeometry* **33**, 153–199 (1991).
30. Leakey, L. *The Stone Age Cultures of Kenya Colony* (Cambridge Univ. Press, 1931).

Acknowledgements

Permission to conduct field work was granted by the Department of Mines of the Ministry of Mines, Energy and Water Resources of Eritrea. We are grateful to A. Kibreab, T. Keleta, A. Mesfin and M. Abraha, G. Ogubazghi and B. Woldehaimanot, and the staff of the National Museum of Eritrea for their support. Financial support was provided by grants from Anadarko Petroleum Company, the National Science Foundation, the Institute of Human Origins, the Royal Ontario Museum and the University of Toronto. Donations by TOTAL Eritrea are gratefully acknowledged. We thank M. Kusmu, L. Le Vert, the French Ministry of Foreign Affairs, the Canadian Consulate and the US Embassy in Asmara. We also thank M. Abdelsalam, T. Andemariam, J. Aronson, B. Collins, C. Cermignani, M. Chazan, S. Churchill, D. Doumenc, E. Goodman, M. Gorton, D. Guinot, Y. Lam, A. Martyn, N. Mohammed, N.-H. Nguyen, J. Pandolfi, K. Reed, H. Shoshani, G. Smithwalter, R. Stern, M. Tesfaye, J. Trondle, M. de Saint Laurent, F. Taylor, G. Watson, C. Vondra and T. Yemane for assistance and/or discussions. Bifaces were first observed on the coast near Abdur by C. Hillman and reported to RCW. in 1995.

Correspondence and requests for materials should be addressed to R.C.W. (e-mail: rwalter@cicese.mx).

The origin of red algae and the evolution of chloroplasts

David Moreira*†, Hervé Le Guyader* & Hervé Philippe*

* Equipe Phylogénie et Evolution Moléculaires, CNRS UPRES-A 8080, Université Paris-Sud, Bâtiment 444, 91405 Orsay Cedex, France
† Departamento de Microbiología, Universidad Miguel Hernández, Campus de San Juan, 03550 Alicante, Spain

Chloroplast structure and genome analyses support the hypothesis that three groups of organisms originated from the primary photosynthetic endosymbiosis between a cyanobacterium and a eukaryotic host: green plants (green algae + land plants), red algae and glaucophytes (for example, *Cyanophora*)¹. Although phylogenies based on several mitochondrial genes support a specific green plants/red algae relationship^{2,3}, the phylogenetic analysis of nucleus-encoded genes yields inconclusive, sometimes contradictory results^{3,4}. To address this problem, we have analysed an alternative nuclear marker, elongation factor 2, and included new red algae and protist sequences. Here we provide significant support for a sisterhood of green plants and red algae. This sisterhood is also significantly supported by a multi-gene analysis of a fusion of 13 nuclear markers (5,171 amino acids). In addition,

the analysis of an alternative fusion of 6 nuclear markers (1,938 amino acids) indicates that glaucophytes may be the closest relatives to the green plants/red algae group. Thus, our study provides evidence from nuclear markers for a single primary endosymbiosis at the origin of these groups, and supports a kingdom Plantae comprising green plants, red algae and glaucophytes⁵.

A reliable phylogenetic framework for both chloroplasts and their hosts is essential to establish the number and nature of primary endosymbioses. Attempts have previously been made using complete chloroplast genome sequences⁶, and several nuclear^{3,7,8} and mitochondrial⁷ markers. The results have been contradictory, especially in what concerns the phylogenetic relationships between two groups of organisms, green plants and red algae, which seem to be derived directly from the primary photosynthetic endosymbiosis. In fact, some markers support their sisterhood^{2,3}, but the RNA polymerase II (*RPB1*) analysis⁴ seems to reject this possibility. Phylogenetic reconstruction artefacts are probably at the origin of these contradictions. Most of these artefacts are due to unequal evolutionary rates among species, such as the well-known long branch attraction (LBA), or the use of inadequate models of sequence evolution^{9,10}. Previous phylogenetic analyses have suggested that elongation factor 2 (EF-2) may be a helpful marker because, for example, its use allows recovering the Microsporidia/Fungi sisterhood¹¹.

To test the relationship of green plants and red algae, we have determined the EF-2 sequences of the rhodophytes *Chondrus crispus* and *Gelidium canariensis*, as well as those of the euglenozoan *Euglena gracilis* and the ciliates *Stylonychia mytilus* and *Tetrahymena pyriformis* to have a more comprehensive data set. In addition, we retrieved the EF-2 sequences of the apicomplexan *Plasmodium falciparum* and the land plant *Triticum aestivum* from continuing genome projects. The resulting data set contained 24 complete eukaryotic EF-2 sequences. As an outgroup, we used either archaeal EF-2 sequences or U5 small nucleolar ribonucleoprotein (U5 snRNP)-specific protein sequences. The U5 snRNP-specific protein is an EF-2 homologue that is integrated in the eukaryotic spliceosome¹². Because it is more similar to eukaryotic EF-2 than

are the archaeal EF-2 sequences, it could help to minimize potential phylogenetic reconstruction artefacts derived from the use of a distant outgroup⁹.

Figure 1 shows the result of an exhaustive maximum-likelihood (ML) search rooted using U5 snRNP-specific protein sequences. It is similar to those produced by distance or maximum-parsimony (MP) analyses (data not shown). Red algae branch with green algae (*Chlorella kessleri*) and land plants (*T. aestivum* and *Beta vulgaris*) within a monophyletic group supported by a 100% bootstrap proportion (BP). The tree consists mainly of two clusters, one encompassing metazoans, fungi and microsporidians, and another including almost all other taxa. The support for this second group is weak (31% BP, data not shown) but, notably, the overall topology of the EF-2 tree fits most of the results obtained using other markers¹³.

We further analysed the reliability of the photosynthetic cluster formed by green plants and red algae (GR) by a Kishino–Hasegawa test¹⁴. The difference of the log-likelihood ($\Delta \ln L$) between the ML tree and the best tree showing the non-monophyly of green plants and red algae was 2.22 s.e. if among-site rate variation (ASRV) was not taken into account, and 3.41 s.e. if a Γ -law was used to correct for ASRV (Table 1). This difference is significant (that is, above 1.96 s.e.), and thus the non-monophyly of red algae and green plants is confidently rejected by the EF-2 data.

Our results differ from previous findings⁴ of significant support ($\Delta \ln L -2.16$ s.e.) for the separation of red algae from a group comprising green plants, fungi and animals, using 13 RNA polymerase *RPB1* sequences rooted with *RPC1* sequences as an outgroup. We have re-analysed the *RPB1* phylogeny taking advantage of the larger taxonomic diversity available today (19 eukaryotic sequences), and using as an outgroup the fused sequences of the two-subunit archaeal homologues, which show a slightly higher similarity to eukaryotic *RPB1* sequences than to *RPC1* sequences. With this enlarged taxonomic sampling and closer outgroup, red algae still branches before green plants in the best *RPB1* ML tree (data not shown). In contrast with the previous analysis⁴, however, this result is no longer significant, as the $\Delta \ln L$ between this tree and the first showing the monophyly of GR is now only -0.79 s.e. without considering ASRV, and -0.28 s.e. using a Γ -law (Table 1).

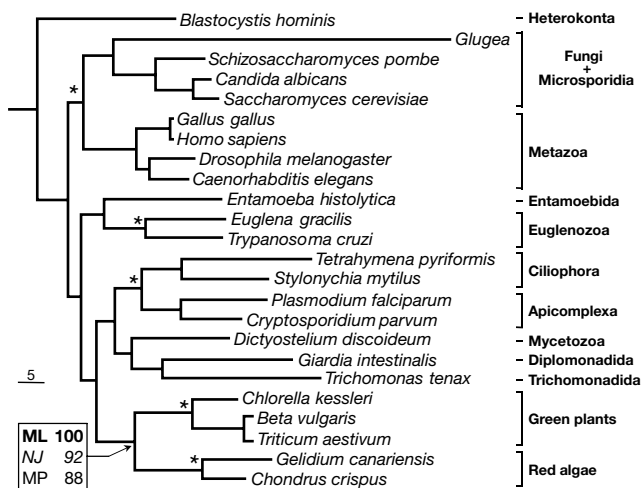


Figure 1 Maximum likelihood tree of eukaryotic EF-2 protein sequences. Maximum-likelihood (ML), distance (NJ) and maximum parsimony (MP) bootstrap proportions supporting the green plants/red algae node are shown. Subtrees marked with asterisks were retrieved with high bootstrap proportions (>90%) in unconstrained ML, NJ and MP analyses, and were constrained in the exhaustive ML analysis to reduce the number of possible trees. *Glugea* refers to the species *Glugea plecoglossi*. The outgroup branch is not shown. Scale bar represents the number of substitutions per 100 sites for a unit branch length.

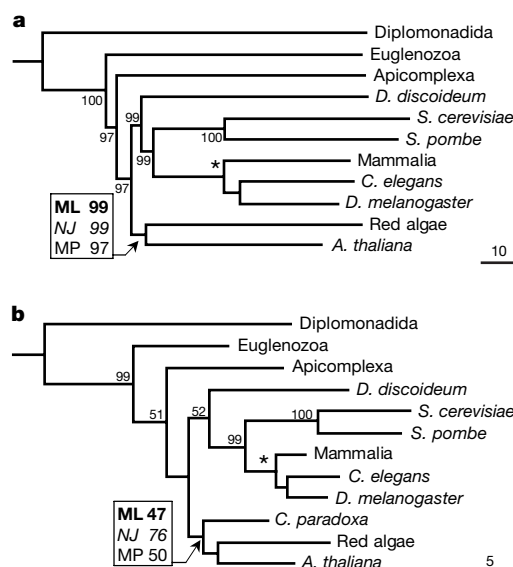


Figure 2 Maximum likelihood tree inferred from exhaustive analysis of fusions F1 (13 genes, 5,171 sites) (a) and F2 (6 genes, 1,938 sites) (b); ML, NJ and MP bootstrap proportions are shown for the green plants/red algae and green plants/red algae/glaucophytes nodes, respectively. Only ML bootstrap proportions are given for the other nodes. Asterisks indicate a constrained group. The outgroup branch is not shown. Scale bar represents the number of substitutions per 100 sites for a unit branch length.

Therefore, the analysis of *RPB1* data does not significantly reject the GR monophyly. Red algae *RPB1* sequences are divergent, indicating that their emergence before the green plants sequences may probably be the result of a LBA artefact. Moreover, the smaller value of $\Delta \ln L$ found using a Γ -law is in agreement with this hypothesis, as correction for ASRV reduces LBA problems⁹.

Although the EF-2 analysis provides significant support for a monophyletic origin of red algae and green plants, we have carried out an additional multi-gene analysis to avoid the biases that often affect single-gene approaches¹⁵. For this purpose, we conducted database searches for all nuclear protein markers with red algae and green plants representatives. We recorded 13 markers, including EF-2 and *RPB1*, with an adequate taxonomic sampling and without problems of paralogy (Table 1). We carried out distance, MP and exhaustive ML phylogenetic analyses, using almost identical taxonomic samplings for all of them. Six of these markers supported the GR monophyly, but the rest supported different relationships for green plants and red algae, without any common pattern (Table 1). Among all the observed topologies, the most represented was the GR monophyly; however, none of these markers gave a significant support under a Kishino–Hasegawa test for either GR monophyly or non-monophyly (Table 1). Because the use of longer sequences can improve phylogenetic reconstruction^{15,16}, we carried out an analysis of the 13 proteins concatenated within a multi-gene fusion (fusion F1, 5,171 amino acids and 12 taxa; Fig. 2a). Although the limited number of taxa could render our analysis sensitive to LBA^{15,16}, the simultaneous use of genes with very different functions makes it unlikely that the same taxa have high evolutionary rates for all of them. The best ML tree showed a high bootstrap value (99% BP) for the GR monophyly and, moreover, it was significantly supported by a Kishino–Hasegawa test with a $\Delta \ln L$ of 2.37 s.e. without assuming ASRV, but was slightly below the significance limit when a Γ -law was used (1.91 s.e.). Nevertheless, the best tree with non-monophyletic GR obtained using a Γ -law does not correspond to an early emergence of red algae, but to the inclusion of Apicomplexa within the GR group. In fact, the best tree showing the emergence of red algae before green plants, fungi, and metazoans (as found previously⁴) displays a significantly worse $\Delta \ln L$ value of –2.99 s.e. under a Γ -law. The early emergence of red algae is therefore also rejected.

Because the GR monophyly can be reliably deduced from our analyses (high BPs for EF-2 and fusion F1, and significant Kishino–

Hasegawa test values in three out of four analyses), the question is now: which are the closest relatives to the GR cluster? The analysis of chloroplast structure and genome suggests that the third eukaryotic group that originated by primary endosymbiosis was the Glaucophyta (with *Cyanophora paradoxa* as the best-known representative)^{1,6,17}. We tested this hypothesis using a new multi-gene fusion (F2, 1,938 amino acids) which included the 6 genes with red algae, green plants and Glaucophyta representatives in databases. Notably, the best tree obtained from an exhaustive ML search shows that red algae, green plants and Glaucophyta make up a monophyletic group (Fig. 2b); however, this group is weakly supported (47% BP) and is not significant under a Kishino–Hasegawa test ($\Delta \ln L$ 0.45 s.e. without considering ASRV; $\Delta \ln L$ 0.34 s.e. using a Γ -law). Therefore, although the available data suggest the sisterhood of these three groups, more sequences are necessary to produce significant results.

Individual genes have proven to be insufficient to determine the relative branching order of the eukaryotic taxa⁷, but the use of 13 concatenate nuclear markers has allowed us to place the red algae robustly. For the first time to our knowledge, the results obtained from the analysis of nuclear-encoded proteins are in clear agreement with those derived from chloroplast and mitochondrial markers, which constitutes a strong piece of evidence for the monophyly of red algae and green plants. Most probably, their closest relatives are glaucophytes, which favours previous claims⁵ for a kingdom Plantae grouping these three taxa. This also has important implications concerning the evolution of chloroplasts. Indeed, the congruence of chloroplast data with our nuclear data shows that a single primary endosymbiosis took place before the separation of red algae and green plants, and, very likely, it even predated the divergence of glaucophytes. The recognition of the GR monophyly allows some inferences about the characteristics of the ancestral plastid, especially if glaucophytes are actually the closest relatives to this group. The ancestral plastid must have contained chlorophylls *a* and *b*, phycobilisomes and unstacked thylakoids. Probably, these ancestral features were differentially lost in the three lineages (chlorophyll *b* in red algae and glaucophytes; phycobilisomes and unstacked thylakoids in green plants). These important changes illustrate the biochemical and structural plasticity of chloroplasts. After the primary endosymbiosis, several secondary endosymbioses (that is, endosymbioses of photosynthetic eukaryotes within non-photosynthetic ones) occurred, for example in heterokonts, dinoflagellates and cryptophytes¹. The existence of a single primary endosymbiosis but several secondary ones suggests that the establishment of this first association between a cyanobacterium and a eukaryotic host (two very distantly related organisms) may have been much more difficult than the subsequent symbioses involving already integrated algae and different eukaryotic hosts. □

Methods

Experimental methods

Genomic DNA from *C. crispus*, *G. canariensis*, *S. mytilus* and *T. pyriformis* was purified, using agarose inclusions¹⁸. *E. gracilis* complementary DNA was prepared using the Enhanced Avian polymerase chain reaction with reverse transcriptase (RT–PCR) kit (Sigma) from total RNA purified with the SV Total RNA Isolation System (Promega). The EF-2 genes were isolated by PCR amplification, using genomic DNA as templates and EF-2-specific primers¹⁹, and sequenced.

Phylogenetic analyses

All available EF-2 and *RPB1* homologous sequences were retrieved from GenBank by BLAST²⁰ searches. All other Rhodophyta nuclear protein-encoding genes existing in GenBank were detected by keyword searches and BLAST searches were conducted for all them. We retrieved 13 genes with an adequate taxonomic representation (Table 1). We also conducted searches upon unfinished complete genome sequencing projects: *Arabidopsis thaliana* (<http://www.kazusa.or.jp/arabi/simsearch.html>); *Dictyostelium discoideum* (<http://www.csm.biol.tsukuba.ac.jp/cDNAproject.html>); *Giardia lamblia* (<http://evolv3.mbl.edu/Giardia-HTML/giardidata.html>); *P. falciparum* (<http://www3.ncbi.nlm.nih.gov/BLAST/unfinishedgenome.html>); *T. aestivum* (<http://www.ti-gr.org/tdb/ogi/>). We aligned EF-2, *RPB1* and the set of 13 proteins in individual files with

Table 1 Results of maximum likelihood analyses

Marker	No. of sites	α -Parameter	GR	$\Delta \ln L$ (s.e.) PROTML	$\Delta \ln L$ (s.e.) PUZZLE
EF-2	612	1.04	Yes	+2.22	+3.41
<i>RPB1</i>	734	0.67	No	–0.79	–0.28
Actin*	365	0.61	Yes	+0.33	+0.45
α -Tubulin*	395	1.02	Yes	+0.03	+0.19
β -Tubulin*	375	0.75	No	–0.11	–1.45
EF-1 α *	416	0.45	No	–0.73	–1.17
GAPDH†‡	265	0.54	No	–0.01	–0.58
Hsp70*	544	0.41	Yes	+0.63	+0.90
Hsp90*	284	1.17	No	–0.40	–0.60
RPS8*	109	0.86	No	–0.52	–0.71
TPI*	203	0.63	No	–0.76	–0.82
vatA†	522	0.46	Yes	+0.50	+0.19
vatB†	435	0.49	Yes	+0.19	+0.06
Fusion F1	5,171	0.55	Yes	+2.37	+1.91
Fusion F2	1,938	0.58	Yes§	+0.45	+0.34

α -Parameter indicates the shape-parameter values of the Γ -law for among-site rate variation. GR corresponds to the green plants/red algae monophyly. $\Delta \ln L$ indicates the values obtained with the Kishino–Hasegawa test for the GR monophyly. Positive and negative values correspond to the cases where the best tree showed monophyletic GR and non-monophyletic GR, respectively. Significant values are given in bold.

* Data sets with reduced taxonomic sampling (12 taxa).

† Cytosolic GAPDH.

‡ Vacuolar ATPase.

§ This case refers to the green plants/red algae/glaucophytes monophyly.

CLUSTALW (ref. 21); the resulting alignments were manually inspected using the ED program from the MUST package²². Only unambiguously aligned positions were used in phylogenetic analyses. A fusion F1 was built for 12 operational taxonomic units (OTUs) (Fig. 2) by concatenation of the 13 proteins (5,171 sites: 1,597 constant, 3,574 polymorphic). A fusion F2 (1,938 sites: 672 constant, 1,266 polymorphic) was constructed to include the glaucophyte *C. paradoxa*. It contained data from 6 proteins (actin, α -tubulin, β -tubulin, EF-1 α , Hsp70 and ATPase vatB) and 13 OTUs (Fig. 2b). Fusion alignments and accession numbers for all sequences used are given in Supplementary Information.

Distance, MP and ML analyses were carried out with all individual and concatenated data sets using MUST (ref. 22), PAUP 3.1 (ref. 23), and PROTML 2.3 (ref. 24), respectively. To carry out exhaustive ML analyses of all possible tree topologies, several constraints were imposed to reduce the number of possible trees (see Figs 1 and 2). Calculation of α -parameter values and other ML analyses taking into account ASRV were conducted using the program PUZZLE (ref. 25). Maximum-likelihood bootstrap proportions were computed using the RELL method²⁶ upon the 5,000 top-ranking trees. For distance and parsimony analyses, 1,000 bootstrap replicates were computed. Kishino–Hasegawa tests¹⁴ were carried out comparing InL of the best 50 trees containing monophyletic GR with InL of the best 50 trees without monophyletic GR, using PROTML 2.3 (ref. 24) (without considering ASRV) and PUZZLE²⁵ (with a Γ -law to correct for ASRV). In the case of fusion F2, we used Kishino–Hasegawa tests to estimate the Δ InL between the best 50 trees containing monophyletic GR + Glaucophyta and the best 50 trees without this monophyletic group.

Received 22 December 1999; accepted 29 February 2000.

- Douglas, S. E. Plastid evolution: origins, diversity, trends. *Curr. Opin. Genet. Dev.* **8**, 655–661 (1998).
- Burger, G., Saint-Louis, D., Gray, M. W. & Lang, B. F. Complete sequence of the mitochondrial DNA of the red alga *Porphyra purpurea*. Cyanobacterial introns and shared ancestry of red and green algae. *Plant Cell* **11**, 1675–1694 (1999).
- Ragan, M. & Guttell, R. Are red algae plants? *Bot. J. Linn. Soc.* **118**, 81–105 (1995).
- Stiller, J. W. & Hall, B. D. The origin of red algae: implications for plastid evolution. *Proc. Natl Acad. Sci. USA* **94**, 4520–4525 (1997).
- Cavalier-Smith, T. Eukaryote kingdoms: seven or nine? *Biosystems* **14**, 461–481 (1981).
- Martin, W. *et al.* Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* **393**, 162–165 (1998).
- Kumar, S. & Rzhetsky, A. Evolutionary relationships of eukaryotic kingdoms. *J. Mol. Evol.* **42**, 183–193 (1996).
- Bhattacharya, D. & Weber, K. The actin gene of the glaucocystophyte *Cyanophora paradoxa*: analysis of the coding region and introns, and an actin phylogeny of eukaryotes. *Curr. Genet.* **31**, 439–446 (1997).
- Philippe, H. & Laurent, J. How good are deep phylogenetic trees? *Curr. Opin. Genet. Dev.* **8**, 616–623 (1998).
- Swofford, D. L., Olsen, G. J., Waddell, P. J. & Hillis, D. M. in *Molecular Systematics* (eds Hillis, D. M., Moritz, C. & Mable, B. K.) 407–514 (Sinauer Associates, Sunderland, Massachusetts, 1996).
- Hirt, R. P. *et al.* Microsporidia are related to fungi: evidence from the largest subunit of RNA polymerase II and other proteins. *Proc. Natl Acad. Sci. USA* **96**, 580–585 (1999).
- Fabrizio, P., Lagerbauer, B., Lauber, J., Lane, W. S. & Luhrmann, R. An evolutionarily conserved U5 snRNP-specific protein is a GTP-binding factor closely related to the ribosomal translocase EF-2. *EMBO J.* **16**, 4092–4106 (1997).
- Embley, T. M. & Hirt, R. P. Early branching eukaryotes? *Curr. Opin. Genet. Dev.* **8**, 624–629 (1998).
- Kishino, H. & Hasegawa, M. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidae. *J. Mol. Evol.* **29**, 170–179 (1989).
- Graybeal, A. Is it better to add taxa or characters to a difficult phylogenetic problem? *Syst. Biol.* **47**, 9–17 (1998).
- Hillis, D. M. Inferring complex phylogenies. *Nature* **383**, 130–131 (1996).
- Bhattacharya, D. & Schmidt, H. A. in *Origins of Algae and Their Plastids* (ed. Bhattacharya, D.) 139–148 (Springer, Vienna, New York, 1997).
- Moreira, D. Efficient removal of PCR inhibitors using agarose-embedded DNA preparations. *Nucleic Acids Res.* **26**, 3309–3310 (1998).
- Yamamoto, A., Hashimoto, T., Asaga, E., Hasegawa, M. & Goto, N. Phylogenetic position of the mitochondrion-lacking protozoan *Trichomonas tenax*, based on amino acid sequences of elongation factors 1 α and 2. *J. Mol. Evol.* **44**, 98–105 (1997).
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
- Philippe, H. MUST, a computer package of management utilities for sequences and trees. *Nucleic Acids Res.* **21**, 5264–5272 (1993).
- Swofford, D. L. PAUP: phylogenetic analysis using parsimony. Version 3.1.1. (Illinois Natural History Survey, Champaign, 1993).
- Adachi, J. & Hasegawa, M. MOLPHY version 2.3: programs for molecular phylogenetics based on maximum likelihood. *Comput. Sci. Monogr.* **28**, 1–150 (1996).
- Strimmer, K. & von Haeseler, A. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* **13**, 964–969 (1996).
- Kishino, H., Miyata, T. & Hasegawa, M. Maximum likelihood inference of protein phylogeny, and the origin of chloroplasts. *J. Mol. Evol.* **31**, 151–160 (1990).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank P. Lopez, P. López-García and M. Müller for critical reading of the manuscript; C. R. Engel for DNA samples; G. Fryd for cultures; N. Narradon for technical help; and the

Kazusa Institute, Marine Biological Laboratory, Sanger Centre, Stanford Centre, the Institute of Genomic Research and Tsukuba Laboratory for access to sequence data. D. M. is the recipient of a stipend from the Fondation des Treilles.

Correspondence and request for materials should be addressed to D. M. (e-mail: david.moreira@bc4.u-psud.fr). EF-2 sequences have been deposited in GenBank under accession numbers AF213661, AF213662, AF213663, AF213664 and AF213665.

Encoding of movement time by populations of cerebellar Purkinje cells

Peter Thier*, Peter W. Dicke*, Roman Haas* & Shabtai Barash†

* Department of Cognitive Neurology, University of Tübingen, Hoppe-Seyler-Straße 3, 72076 Tübingen, Germany

† Department of Neurobiology, The Weizmann Institute, 76100 Rehovot, Israel

One of the earliest computational principles attributed to the cerebellum was the measurement of time¹. This idea was originally suggested on anatomical grounds, and was taken up again to explain some of the deficits in cerebellar patients^{2,3}. The contribution of the cerebellum to eye movements, in contrast, has traditionally been discussed in the context of motor learning^{4–7}. This view has received support from the loss of saccade adaptation, one of the key examples of motor learning, following lesions of the posterior cerebellar vermis^{8–11}. However, the relationship between the properties of saccade-related vermal Purkinje cells and the behavioural deficits that follow lesions is unclear. Here we report results from single-unit recording experiments on monkeys that reconcile the seemingly unrelated concepts of timing and motor learning. We report that, unlike individual Purkinje cells, the population response of larger groups of Purkinje cells gives a precise temporal signature of saccade onset and offset. Thus a vermal population response may help to determine saccade duration. Modifying the time course of the population response by changing the weights of the contributing individual Purkinje cells, discharging at different times relative to the saccade, would directly translate into changes in saccade amplitude.

We recorded extracellularly from 131 saccade-related Purkinje cells (PCs) in vermal lobuli VI and VIIA of two rhesus monkeys, while the animals performed visually guided saccades. The preferred direction was determined by asking the monkeys to make 15° centre-out saccades in eight directions. As in previous work on saccade-related responses in the oculomotor vermis^{12–15}, these saccade-related responses were characterized by bursts of activity in tight temporal relationship with the execution of a saccade in particular preferred directions (Fig. 1). Only a few PCs (5/131) with saccade-related bursts lacked this directional preference. To ensure that the saccade-related bursts did not reflect a visual response evoked by the peripheral target, we tested 98 of the PCs considered on a 'memory saccade' task, separating the presentation of the peripheral target and the saccade in time. The monkeys were asked to saccade to a remembered location in the frontoparallel plane, cued by the brief presentation of a peripheral target. Of the 98 PCs tested, 71 lacked a significant visual response to the presentation of the peripheral cue in this memory saccade task and only five showed a visual response reaching 50% or more of the later saccade-related burst.

The tight alignment of saccade-related bursts exhibited by many of the PCs with the saccades indicated that vermal PCs might help to

determine the timing of saccades. We therefore investigated whether single PC bursts predict saccade duration. To obtain saccades with sufficiently variable durations for each neuron, we exploited the fact that saccade duration increases monotonically with saccade amplitude¹⁶ (Fig. 2a). We collected responses to saccades of different amplitudes, from 5° to 40°, all in a PC's best direction (or alternatively a fixed direction for non-directional cells). Figure 1 shows the responses of four different PCs, exemplifying the different dependencies of responses on saccade amplitude exhibited by vermal PCs. Some vermal PCs (Fig. 1a) showed clear preferences for specific saccade amplitudes (about 20° in this example) with increasingly weaker responses for amplitudes deviating increasingly from the optimum. Other PCs (Fig. 1b, c) displayed no or little influence of saccade amplitude. Still others (Fig. 1d) showed a monotonic dependence of response on saccade amplitude. To characterize the dependence of saccade-related bursts of vermal PCs on saccade amplitude or duration, we computed several parameters characterizing the saccade-related bursts of each cell.

Figure 2 shows the means and standard errors of the parameters considered for the 94 PCs for which amplitude-tuning data was available. The mean burst onset latency, relative to saccade onset, did not depend on saccade duration (Fig. 2b). The burst offset latency increased with saccade duration (Fig. 2c). Consequently, the

mean burst duration increased with saccade duration (Fig. 2d). As the mean discharge rate in the burst did not depend significantly on saccade duration (Fig. 2e), the increase in burst duration with saccade duration must fully account for the modest increase in the mean number of spikes (Fig. 2f). These results may indicate that the timing of the PC bursts reflects the timing of the saccade: certainly, the mean firing rate, computed over the whole burst, carries no information (Fig. 2e). However, the variability of the data, conveyed by the very large errors of the means, confounds the interpretation of Fig. 2. This variability is likely to be caused, in part, by the procedure of first deriving parameters for each PC and then analysing these parameters rather than deriving parameters from the whole sample of responses in one step.

Any biologically plausible mechanism involved in extracting information from a larger group of PCs is likely to be sensitive to the timing of individual bursts and even the timing of individual spikes. We therefore investigated whether the collective instantaneous discharge rate of all PCs in the sample might reflect more faithfully saccade duration or other aspects of saccade timing. Figure 3a shows the collective saccade-related instantaneous discharge rate, henceforth referred to as the population burst, as a function of saccade duration. It was computed by sorting the saccade-related responses of the cells whose responses were

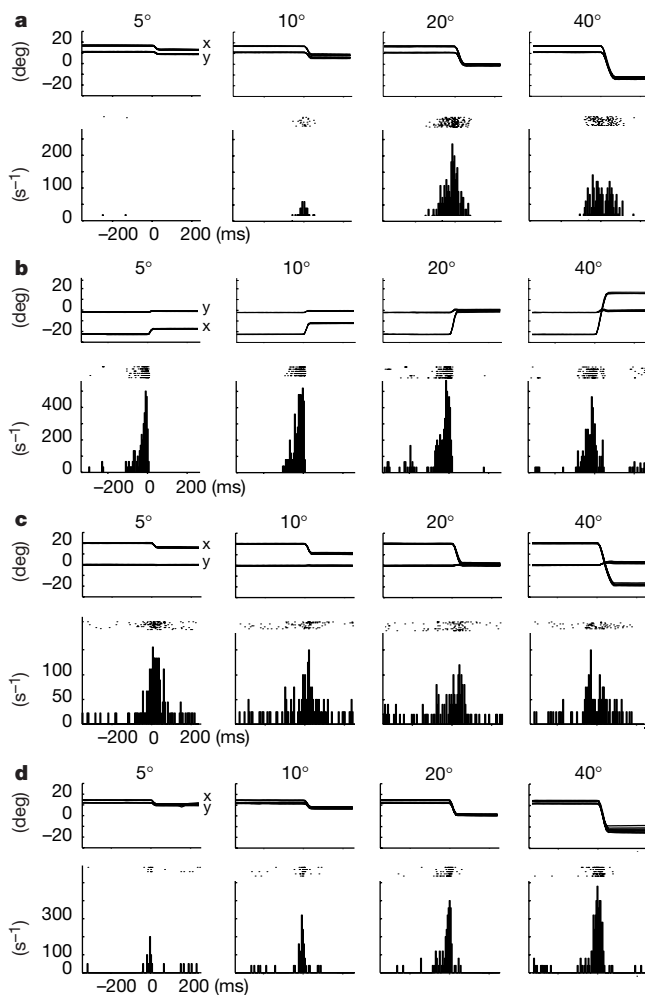


Figure 1 Perisaccadic time histograms and raster plots obtained from vermal PCs, reflecting the highly variable dependence of saccade-related bursts on saccade amplitude. **a–d**, Four individual PCs. *x*, *y*: horizontal and vertical components of eye position. Width of time bins, 5 ms.

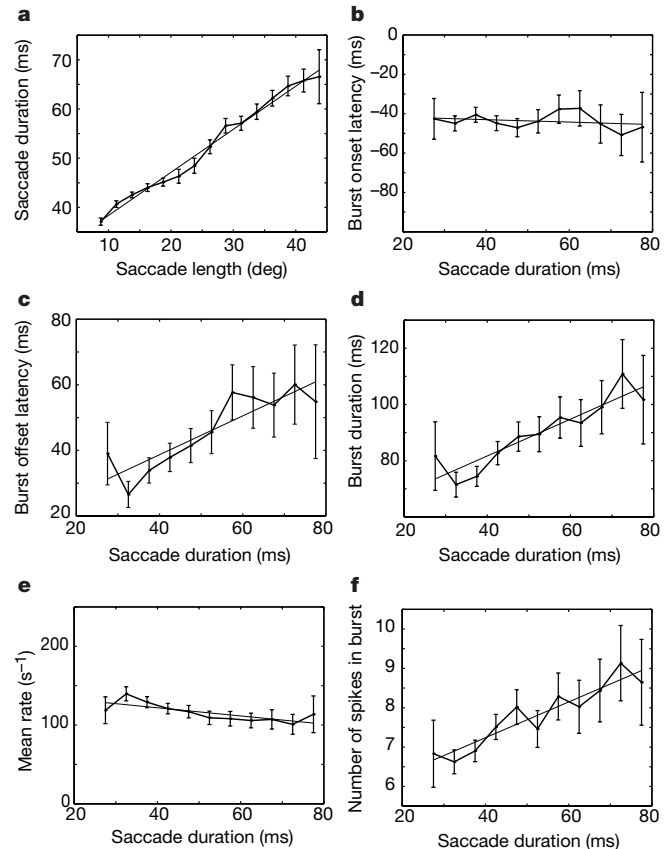


Figure 2 Dependence of saccade duration on saccade length (**a**) and burst parameters on saccade durations (**b–f**). **a**, Mean saccade duration as a function of saccade length, based on eye movement records collected with the 94 single units whose amplitude tuning was studied. Vertical bars, s.e. Saccade duration increases linearly with saccade amplitude (thin line; slope $m = 0.875$; $P < 10^{-12}$). **b–f**, Dependence of various features of saccade-related bursts of vermal PCs on saccade duration. The plots show mean $\pm$ s.e. The three variables that showed a significant (linear regression (thin lines), $P < 0.05$) but modest dependence on saccade duration were burst offset latency (**c**, $P < 0.00031$), burst duration (**d**, $P < 0.0077$) and the overall number of spikes in the burst (**f**, $P < 0.0031$).

characterized in Fig. 2, according to saccade duration (bin width 2.5 ms). Next, for each saccade duration, we computed compound perisaccadic histograms with a bin-width of 1 ms, smoothed by a gaussian filter with a standard deviation of 10 ms. These compound perisaccadic histograms were then plotted as function of saccade duration, aligned relative to saccade onset for each of the eight directions tested, and the mean of the plots for the individual directions was taken as the estimate of the population burst.

The population burst is shown in Fig. 3a as it develops over time (x-axis) for saccades of duration 30–65 ms (y-axis). The upper part of Fig. 3b displays three individual sections through the colour-coded plot shown in Fig. 3a, illustrating population bursts for saccades of 30, 49 and 65 ms, respectively. The population response, independent of saccade duration, starts to rise above the baseline level before saccade onset, peaks at saccade onset, and then declines again and reaches the baseline level when the saccade ends (Fig. 3a, b).

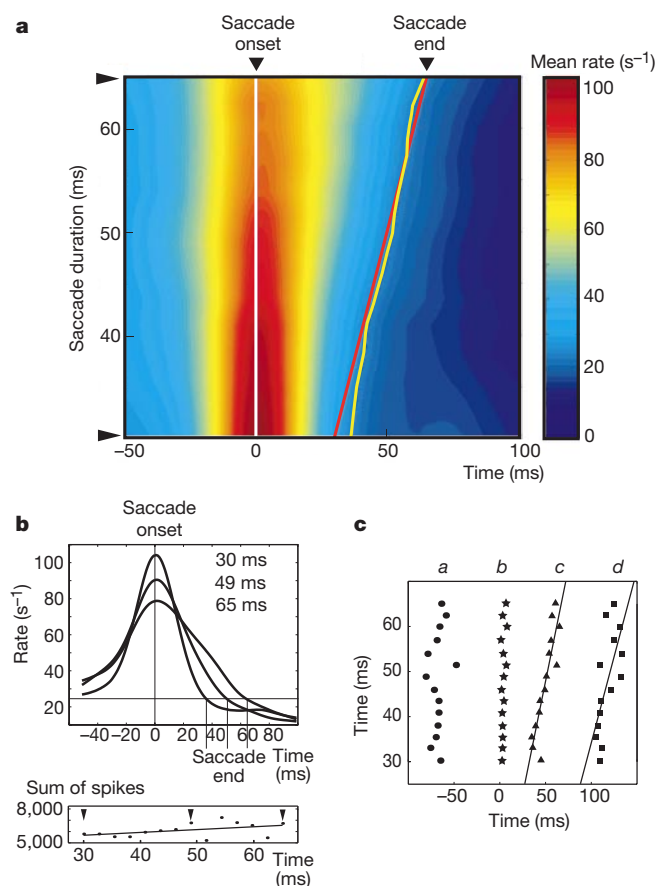


Figure 3 Dependence of the population burst on saccade duration. **a**, Population burst (94 vermal saccade-related PCs) plotted as a function of saccade time (x-axis; 0, saccade onset) and duration (y-axis), measured in discrete time steps (x-axis: 1 ms, y-axis: 2.5 ms). We used a gaussian filter (s.d. 10 ms) to smooth the plot along the x-axis and a Savitzky–Golay filter²⁹ (window = 3; deg = 1) along the y-axis. Yellow line, population activity four times the mean baseline activity (6 spikes s⁻¹; see Methods). **b**, Top, population burst profiles represent sections through the three-dimensional plot shown in **a** at individual saccade durations of 30, 49 and 65 ms. Bottom, linear regression on the overall number of spikes in the population burst, as a function of saccade duration, revealed no significant dependence ($P > 0.05$). **c**, Plots of time t as a function of: *a*, population burst onset latency; *b*, time to peak; *c*, time at end; and *d*, population burst duration ($c-a$). Burst onset and burst end are defined by a $4 \times$ baseline criterion (see text). In the case of *a*, *b* and *d*, t is saccade duration; in the case of *c*, t is the time of saccade termination. Both *c* and *d* increase significantly with t ($t = 1.0138c - 3.9389$, $P < 0.0028$; $t = 0.7636d - 42.333$, $P < 0.01$; the slopes for *c* and *d* are significantly different, $P < 0.031$), whereas neither *a* nor *b* depends on saccade duration ($P > 0.05$).

To relate the time course of the population burst more precisely to the time course of the saccade, we determined the times of onset (*a*), peak (*b*) and offset (*c*) of the population burst relative to saccade onset for each saccade duration as well as the population burst duration (*d*), defined as $c-a$. We defined population burst onset and offset times as the times when the population burst reached four times the baseline firing rate when building up and when declining. Figure 3c plots time t as function of *a*, *b*, *c* and *d*, respectively. In the case of *a*, *b* and *d*, t corresponds to saccade duration, in the case of *c* to the time of saccade termination. The plots were fitted by linear regressions. Both *c* and *d* increased linearly (c : $P < 0.00003$, d : $P < 0.01$) with the time of saccade termination and saccade duration, respectively, whereas neither *a* nor *b* depended significantly ($P > 0.05$) on saccade duration. The population burst onset leads the saccade onset by 67.9 ± 8.25 ms, whereas the population burst peak appears roughly at the time of saccade onset (2.78 ± 1.97 ms), both independent of saccade duration. The end of the population burst (*c*), as predicted by the regression, corresponds very closely to the end of the saccade, whereas the population burst duration (*d*) underestimates saccade duration by 24% (Fig. 3c). This and the significantly higher coefficient of correlation ($P < 0.004$) for *c* compared with *d* indicate that the population burst reflects the time of saccade termination rather than saccade duration.

Although the peak of the population burst would be too late to be causally related to saccade onset, its precise alignment with saccade onset might be understood as reflecting the zeroing of the 'stop-watch', determining the termination of the saccade. In accordance with this model, the interval $e = c-b$ between the population burst peak *b* and the population burst offset *c* gives a fairly accurate account of saccade duration t (linear regression: $t = 1.121e - 5.126$; $P < 0.0046$). The population burst peak is characterized by a second conspicuous feature, a decrease in its peak amplitude *pa* with saccade duration t (linear regression, $pa = -0.81t + 130.751$; $P < 0.0244$; Fig. 3b). This decrease accompanies an increase in the width of the population burst with increasing saccade duration. These two dependencies combine to leave the overall number of spikes in the population burst largely independent of saccade duration (Fig. 3b, lower). Thus, the decrease in peak firing rate with saccade duration might simply reflect the need to redistribute an approximately constant number of spikes over varying segments of time. Alternatively,

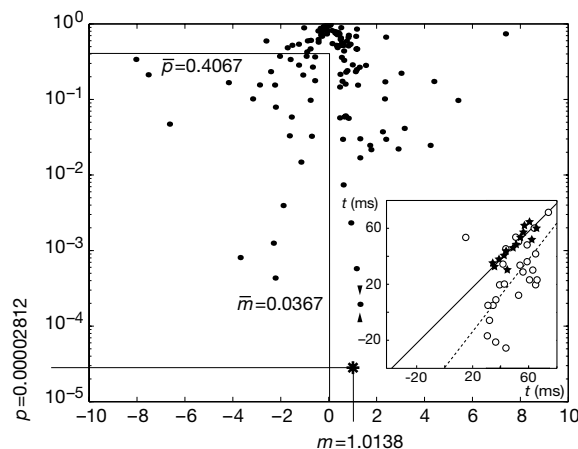


Figure 4 Scatter-plot of p , the probability of the linear regression, fitting plots of saccade end as function of burst end, as a function of its slope m . The dots give the p , m coordinates of individual PCs and the asterisk indicates the p , m coordinate of the population response. The inset compares the regression analyses for the population burst and the 'best' PC (marked by two arrowheads), singled out by the smallest p and an m close to the ideal of 1. The x-axis plots time with $t = 0$ corresponding to saccade onset. The y-axis plots the time of the burst end relative to saccade onset.

the peak firing rate might encode dynamic saccade parameters, which, like saccade velocity, depend on saccade duration¹⁶. Such a putative dependence between peak firing rate and saccade velocity would be specific to the population burst, not shared by individual cells. Both the peak discharge rates and the mean discharge rates in the saccade-related bursts are on average independent of saccade velocity (linear regression analysis, $P > 0.05$). Notwithstanding the question of whether the amplitude of the peak population burst has a function, the timing of the burst is independent of and cannot be explained by the amplitude of the peak population burst.

The faithful description of saccade timing provided by the population burst could reflect the contribution of a subgroup of 'best' PCs or, alternatively, emerge as a true *de novo* feature of the population response. To distinguish between these possibilities, we compared the linear regression for the plot of burst end (c) based on the population burst as shown in Fig. 3c with the corresponding linear regressions for individual PCs. Figure 4 shows the significance level p of the linear regressions as a function of their slopes m . An m of 1 would reflect a faithful representation of saccade end. The slope of the regression plotting saccade end as a function of the population burst end comes very close to this ideal ($m = 1.0138$). On the other hand, the m for the individual PCs is on average close to zero and varies widely ($m_{\text{mean}} = 0.0367 \pm 2.0054$). Moreover, even the best of the individual regressions is not nearly as good, as measured by their m and p coordinates, as the one for the population burst. This is shown in the inset, which compares the regression for the population burst with the regression for the 'best' cell (singled out by arrows in the $p(m)$ plot). In summary, this analysis indicates that the population burst reflects the timing of saccade termination and that this precise reflection is a *de novo* property of the population response, rather than a reflection of the properties of the 'best' PCs.

Does cerebellar population encoding for saccades do more than alleviate the insufficiencies of individual cells? We propose that it could make an important contribution to saccadic plasticity. If we assume that the end of the population burst determines the end of the saccade, a simple way to make a too-small saccade normometric would be to increase the duration of the population response. The necessary increase in the duration of the population response could be achieved by increasing the gain of relatively 'late' vermal PCs, based on an appropriate error signal. There are three findings that support this hypothesis. First, adaptive modification of saccades in humans leads to a selective activation of those parts of the human cerebellum that are probably homologous to the monkey oculomotor vermis¹⁷. Second, the capacity for rapid saccadic plasticity is irreversibly lost following lesions of the posterior vermis in monkeys¹⁰. And finally, some saccade-related neurons in the caudal fastigial nucleus (cFN), which are controlled by vermal PCs^{18–20}, show bursts that occur later if saccades are modified so as to become larger for a given retinal vector²¹. A putative function of this type of cFN burst is the braking of ongoing saccades, and a likely mechanism underlying these bursts is release from PC inhibition, giving rise to a rebound depolarisation^{22,23}. Hence, learning to make larger saccades would involve a sequence of prolonged vermal PC inhibition and a delayed fastigial rebound, ultimately translating into a delayed braking signal for saccades.

Obviously, this scheme, which tries to lead saccadic learning back to an optimization of a representation of time, is not confined to the saccadic system. Rather, it might be applicable to any motor or non-motor function that is dependent on the availability of a precise representation of time. □

Methods

We recorded extracellularly from lobuli VI and VIIA of the posterior vermis of one female and one male rhesus monkey (P and S) while they performed visually guided saccades, evoked by asking the monkeys to saccade to peripheral cues that appeared when the fixation spot disappeared. The monkeys were prepared for chronic single unit and eye position recordings, the latter based on the search coil technique, as described^{24,25}. All animal procedures followed the guidelines set by the NIH and national law and were

approved by the local committee supervising the handling of experimental animals. Recording site localization in lobuli VI and VIIA relied on stereotactic calculations based on postsurgical MRI scans, which showed both these brainstem structures as well as the precise location and orientation of the implanted recording chamber. A conventional reconstruction of recording sites based on post-mortem brain sections was carried out in monkey P. In perfect correspondence with the tentative localization based on MRI, recording sites were confined to lobuli VI and VIIA. Well isolated single units encountered in lobuli VI and VIIA were interpreted as reflecting PCs, without first trying to identify complex spikes. This simplified approach seemed acceptable in view of our previous experience²⁶: using the same type of microelectrode, we found that the vermal single units analysed ($n = 56$) all exhibited complex spikes, identifying them as Purkinje cells. The preferred direction of a neuron was studied by measuring responses to visually guided saccades of 15° or 20° amplitude in eight equally spaced directions (0°, 45° and so on) in the frontoparallel plane, starting from straight ahead. A response was considered to be directionally selective if the mean number of spikes in the best direction was at least twice the response in the worst direction. To reveal visual contamination of saccade-related responses, we next ran a standard centre-out memory saccade task. A location in the frontoparallel plane was cued by the presentation of a peripheral target 600–800 ms after the start of the trial, while the monkey fixated a central spot. The disappearance of the central fixation spot, 700 ms after the disappearance of the peripheral cue, served as the go signal for the saccade. Target locations of memory guided saccades corresponded to those of visually guided saccades.

The visual response to the peripheral target was estimated by calculating the mean discharge rate in a window from 670 to 800 ms. The visual response was considered to be significant if it exceeded the baseline discharge rate plus two s.d. Finally, the amplitude or duration tuning of a PC was determined by asking the monkey to make saccades of varying amplitudes from 2.5° (or 5° in some cases) to 40° in steps of 2.5–5° in the neuron's preferred direction. Amplitudes were presented randomly interleaved. As the field of view was limited to 50° × 45° by the monitor on which stimuli were presented, the starting point of saccade vectors larger than 20° had to be moved to eccentric locations, in a direction opposite to saccade directions. Starting saccades from different orbital positions was allowed, as our own work suggests that, contrary to earlier reports, the influence of the orbital starting position on saccade-related responses of posterior vermal PCs is indeed negligible^{26,27}. The start and end of saccade-related PC bursts were determined in single trials by applying a Poisson spike train analysis²⁸ with a confidence level of usually $P < 0.05$ ($P < 0.1$ in a few cases). To exclude spurious bursts, we used the following qualifications: burst onset had to follow the extinction of the fixation point; and burst onset had to be within a period starting 150 ms before saccade onset and lasting until the end of the saccade. Additional bursts following the first one were detected in case the first burst ended before the end of the saccade. The baseline activity rate was estimated as the mean discharge rate from –400 to –100 ms (memory saccades) or –300 to –100 ms (visually guided saccades) relative to the onset of the peripheral target.

Received 7 December 1999; accepted 18 February 2000.

1. Braitenberg, V. Functional interpretation of cerebellar histology. *Nature* **190**, 539–640 (1961).
2. Ivry, R. B. & Diener, H. C. Impaired velocity perception in patients with lesions of the cerebellum. *J. Cogn. Neurosci.* **3**, 355–366 (1991).
3. Ivry, R. B. & Keele, S. W. Timing functions of the cerebellum. *J. Cogn. Neurosci.* **1**, 136–152 (1993).
4. Marr, D. A theory of cerebellar cortex. *J. Physiol.* **202**, 437–470 (1969).
5. Ito, M. Cerebellar control of the vestibulo-ocular reflex: around the flocculus hypothesis. *Annu. Rev. Neurosci.* **5**, 275–296 (1982).
6. Kawato, M. & Gomi, H. The cerebellum and VOR/OKR learning models. *Trends Neurosci.* **15**, 445–453 (1992).
7. Raymond, J. L., Lisberger, S. G. & Mauk, M. D. The cerebellum: A neuronal learning machine? *Science* **272**, 1126–1131 (1996).
8. Optican, L. M. & Robinson, D. A. Cerebellar-dependent adaptive control of primate saccadic system. *J. Neurophysiol.* **44**, 1058–1076 (1980).
9. Takagi, M., Zee, D. S. & Tamargo, R. J. Effects of lesions of the oculomotor vermis on eye movement in primate: saccades. *J. Neurophysiol.* **80**, 1911–1931 (1988).
10. Barash, S. et al. Saccadic dysmetria and adaptation after lesions of the cerebellar cortex. *J. Neurosci.* **19**, 10931–10939 (1999).
11. Fitzgibbon, E. J. & Goldberg, R. A. in *Adaptive Processes in Visual and Oculomotor Systems* (eds Keller, B. L. & Zee, D. S.) 329–333 (Pergamon, Oxford, 1986).
12. Llinás, R. & Wolfe, J. W. Functional linkage between the electrical activity in the vermal cerebellar cortex and saccadic eye movements. *Exp. Brain Res.* **29**, 1–14 (1977).
13. Kase, M., Miller, D. C. & Noda, H. Discharges of Purkinje cells and mossy fibers in the cerebellar vermis of the monkey during saccadic eye movements and fixation. *J. Physiol. (Lond.)* **300**, 539–555 (1980).
14. Sato, H. & Noda, H. Posterior vermal Purkinje cells in macaques responding during saccades, smooth pursuit, chair rotation and/or optokinetic stimulation. *Neurosci. Res.* **12**, 583–595 (1992).
15. Helmchen, C. & Büttner, U. Saccade-related Purkinje cell activity in the oculomotor vermis during spontaneous eye movements in light and darkness. *Exp. Brain Res.* **103**, 198–208 (1995).
16. Bahill, A. T., Clark, S. A. & Stark, R. J. The main sequence, a tool for studying human eye movements. *Math. Biosci.* **24**, 287–298 (1975).
17. Desmurget, M. et al. Functional anatomy of saccadic adaptation in humans. *Nature Neurosci.* **1**, 524–528 (1998).
18. Ohtsuka, K. & Noda, H. Saccadic burst neurons in the oculomotor region of the fastigial nucleus of macaque monkey. *J. Neurophysiol.* **65**, 1422–1434 (1991).
19. Fuchs, A. F., Robinson, F. R. & Straube, A. Role of caudal fastigial nucleus in saccade generation. I. Neuronal discharge patterns. *J. Neurophysiol.* **70**, 1723–1740 (1993).
20. Noda, H., Sugita, S. & Ikeda, Y. Afferent and efferent connections of the oculomotor region of the fastigial nucleus in the macaque monkey. *J. Comp. Neurol.* **302**, 330–348 (1990).
21. Scudder, C. A. Discharge of fastigial nucleus neurons is altered during adaptive modification of

- saccade size. *Soc. Neurosci. Abstr.* **24**, 147 (1998).
22. Llinás, R. & Mühlethaler, M. Electrophysiology of guinea-pig cerebellar nuclear cells in the *in vitro* brain stem–cerebellar preparation. *J. Physiol. (Lond.)* **404**, 241–258 (1988).
23. Aizenmann, C. D. & Linden, D. J. Regulation of the rebound depolarization and spontaneous firing patterns of deep nuclear neurons in slices of rat cerebellum. *J. Neurophysiol.* **82**, 1697–1709 (1999).
24. Judge, S. J., Richmond, B. J. & Chu, F. C. Implantation of magnetic search coils for measurement of eye position: an improved method. *Vision Res.* **20**, 535–538 (1980).
25. Thier, P. & Erickson, R. G. Responses of visual-tracking neurons from cortical area MSTl to visual, eye and head motion. *Eur. J. Neurosci.* **4**, 539–553 (1992).
26. Thielert, C.-D. *Elektrophysiologische und Anatomische Untersuchungen zum Okulomotorischen Beitrag des Posterioren Vermis des Rhesusaffen*. Thesis, Eberhard-Karls-Universität Tübingen (1996).
27. Haas, R., Dicke, P. W. & Thier, P. Saccade-related responses of most posterior vermal Purkinje cells do not depend on the starting position of the eyes. *Soc. Neurosci. Abstr.* **25**, 1652 (1999).
28. Hanes, D. P., Thompson, K. G. & Schall, J. D. Relationship of presaccadic activity in frontal eye field and supplementary eye field to saccade initiation in macaque: Poisson spike train analysis. *Exp. Brain Res.* **103**, 85–96 (1995).
29. Press, H. W., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (eds) *Numerical Recipes in C*, 650–655 (Cambridge Univ. Press, Cambridge, 1992).

Acknowledgements

This work was funded by the German-Israeli-Foundation and the German Research Council (Forschungsgruppe 'Wahrnehmen und Agieren im Raum'). We thank M. Erb and W. Grodd for their help with the anatomic MRI scans and C. Schwarz and F. Sultan for helpful comments on an earlier version of the manuscript.

Correspondence and requests for materials should be addressed to P.T. (e-mail: thier@uni-tuebingen.de).

Silberblick/Wnt11 mediates convergent extension movements during zebrafish gastrulation

Carl-Philipp Heisenberg^{*,†}, Masazumi Tada^{†,‡}, Gerd-Jörg Rauch[§], Leonor Saúde[‡], Miguel L. Concha^{*}, Robert Geisler[§], Derek L. Stemple[‡], James C. Smith[‡] & Stephen W. Wilson^{*}

^{*} Department of Anatomy and Developmental Biology, University College London, Gower Street, London WC1E 6BT, UK

[‡] Division of Developmental Biology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

[§] Abteilung Genetik, Max-Planck-Institut für Entwicklungsbiologie, Spemannstrasse 35, D-72076 Tübingen, Germany

[†] These authors contributed equally to this work

Vertebrate gastrulation involves the specification and coordinated movement of large populations of cells that give rise to the ectodermal, mesodermal and endodermal germ layers. Although many of the genes involved in the specification of cell identity during this process have been identified, little is known of the genes that coordinate cell movement. Here we show that the zebrafish *silberblick* (*slb*) locus¹ encodes Wnt11 and that Slb/Wnt11 activity is required for cells to undergo correct convergent

extension movements during gastrulation. In the absence of Slb/Wnt11 function, abnormal extension of axial tissue results in cyclopia and other midline defects in the head². The requirement for Slb/Wnt11 is cell non-autonomous, and our results indicate that the correct extension of axial tissue is at least partly dependent on medio-lateral cell intercalation in paraxial tissue. We also show that the *slb* phenotype is rescued by a truncated form of Dishevelled that does not signal through the canonical Wnt pathway³, suggesting that, as in flies⁴, Wnt signalling might mediate morphogenetic events through a divergent signal transduction cascade. Our results provide genetic and experimental evidence that Wnt activity in lateral tissues has a crucial role in driving the convergent extension movements underlying vertebrate gastrulation.

Vertebrate gastrulation is driven by coordinated morphogenetic movements of the three germ layers: ectoderm, mesoderm and endoderm. During this process, convergent extension movements within the mesoderm are generated by polarization of mesodermal cells followed by medio-lateral cell intercalation, which causes cells to become distributed along the antero-posterior axis^{5,6}. Several mutants exhibiting defective gastrulation movements have been identified in zebrafish^{7,8}. In *slb* mutants, extension of the axial mesoderm and overlying ventral central nervous system is disturbed, resulting in fusion of the eyes later in development². As previous results have suggested that Wnt activity and especially Wnt11 might affect the movement of cells during gastrulation^{3,9–11}, it seemed possible that mutations in Wnt11 might underlie the *slb* phenotype. Mapping of *slb* and *wnt11* showed both to be located at a similar position on linkage group 5 (Fig. 1a). Sequence analysis of *slb*^{tx226} and *slb*^{tz216} alleles revealed point mutations at positions Trp 234 and Gly 155, introducing in-frame stop codons that result in truncation of the Wnt11 protein (Fig. 1b). These mutations would be expected to generate non-functional proteins (see below), so both alleles are likely to be null mutations.

Expression analysis of *wnt11* in wild-type and *slb*^{−/−} embryos confirmed that Wnt11 is a good candidate for mediating Slb/Wnt11 activity. In wild-type embryos, *wnt11* expression is first detectable in the germring at the shield stage and subsequently in the paraxial head mesoderm and in the neuroectoderm (Fig. 1c–e and ref. 12). By late gastrulation, the paraxial head mesodermal domain lies directly anterior to presumptive paraxial somitic mesoderm and the lateral neuroectodermal domain is posterior to the presumptive forebrain (Fig. 1i–k and data not shown). In *slb* embryos, *wnt11* expression is initiated but is strongly downregulated in all expression domains by late gastrulation (Fig. 1f–h).

To confirm that the *slb*^{−/−} phenotype is due to mutations in the *wnt11* gene, we overexpressed *wnt11* RNA in mutant and wild-type embryos. In uninjected *slb*^{−/−} embryos, the prechordal plate is more elongated and the notochord is shorter and wider than in wild-type embryos, correlating with a partial fusion of the eyes at later developmental stages² (Fig. 2a, b, g, h). The injection of 10 pg of *wnt11* RNA into wild-type embryos at the one-cell stage had either little or no obvious effect, or caused defective morphogenesis of the

Table 2 *wnt11* and *dsh-ΔN* rescue the *slb* convergent extension phenotype at the tailbud stage

RNA injected	Genotype	Wild type (%)	<i>slb</i> (%)	Abnormal (%)	Total (n)
–	wild type	100	0	0	46
<i>wnt11</i>	wild type	74	0	26	43
<i>wnt11</i> (<i>slb</i> ^{tx226})	wild type	100	0	0	41
–	<i>slb</i>	4	96	0	90
<i>wnt11</i>	<i>slb</i>	36	7	57	41
<i>dsh-ΔN</i>	<i>slb</i>	32	20	48	46
<i>dsh-DEP+</i>	wild type	0	40	60	54

10 pg of *wnt11* and *wnt11* (*slb*^{tx226}) RNA and 200 pg of *dsh-ΔN* and *dsh-DEP+* RNA were injected into one-cell-stage embryos. Embryos classified as '*slb*' displayed an elongated, posteriorly displaced prechordal plate and shortened notochord, and embryos classified as 'abnormal' showed other variable defects in morphogenesis of prechordal plate and notochord.

Table 1 *wnt11* and *dsh-ΔN* rescue the *slb* eye phenotype at pharyngula stage

RNA injected	Genotype	Wild type (%)	<i>slb</i> (%)	Reduced (%)	Total (n)
–	wild type	100	0	0	73
<i>wnt11</i>	wild type	66	0	34	69
<i>wnt11</i> (<i>slb</i> ^{tx226})	wild type	100	0	0	30
–	<i>slb</i>	9	91	0	83
<i>wnt11</i>	<i>slb</i>	46	9	45	94
<i>dsh-ΔN</i>	<i>slb</i>	49	48	3	42
<i>dsh-DEP+</i>	wild type	5	45	50	37

10 pg of *wnt11* and *wnt11* (*slb*^{tx226}) RNA and 200 pg of *dsh-ΔN* and *dsh-DEP+* RNA were injected into one-cell-stage embryos. Eyes classified as '*slb*' showed some degree of cyclopia, and eyes classified as 'reduced' were reduced in size.

prechordal plate and reduced size of the eyes (Fig. 2c, i and Tables 1 and 2). However, overexpression of *wnt11* in *slb*^{-/-} embryos frequently rescued the *slb*^{-/-} phenotype as judged by normal elongation of midline tissue and correct spacing of the eyes, in addition to causing similar defects to those observed in wild-type embryos (Fig. 2d, j and Tables 1 and 2). In contrast, the injection of *slb*^{tx226} RNA into wild-type embryos did not lead to any obvious phenotype (Tables 1 and 2 and data not shown). These results indicate that mutations in *wnt11* are responsible for the *slb*^{-/-} phenotype and that the *slb*^{tx226} allele probably represents a loss-of-function mutation rather than a dominant-negative form, in contrast with other more carboxy-terminally truncated versions of Wnt11 (ref. 3).

To elucidate the intracellular pathway by which Slb/Wnt11 regulates convergent extension movements, we first attempted to rescue the *slb*^{-/-} phenotype by injection of *dsh* RNA, an intracellular mediator of the canonical Wnt signalling pathway¹³. However, ubiquitous overexpression of *dsh* RNA in *slb* mutants dorsalized the embryos and led to a reduction in size of the prechordal plate (data not shown), presumably reflecting the widespread activation of signalling normally mediated by Wnts other than Wnt11^{14,15}. Similarly, the injection of an activated form of β -catenin, ΔN - β -catenin, which is also believed to activate canonical Wnt signalling^{16,17}, dorsalized and did not rescue *slb*^{-/-} embryos, whereas the injection of a dominant-negative form of *Tcf-3*, ΔN -*Tcf3*, which

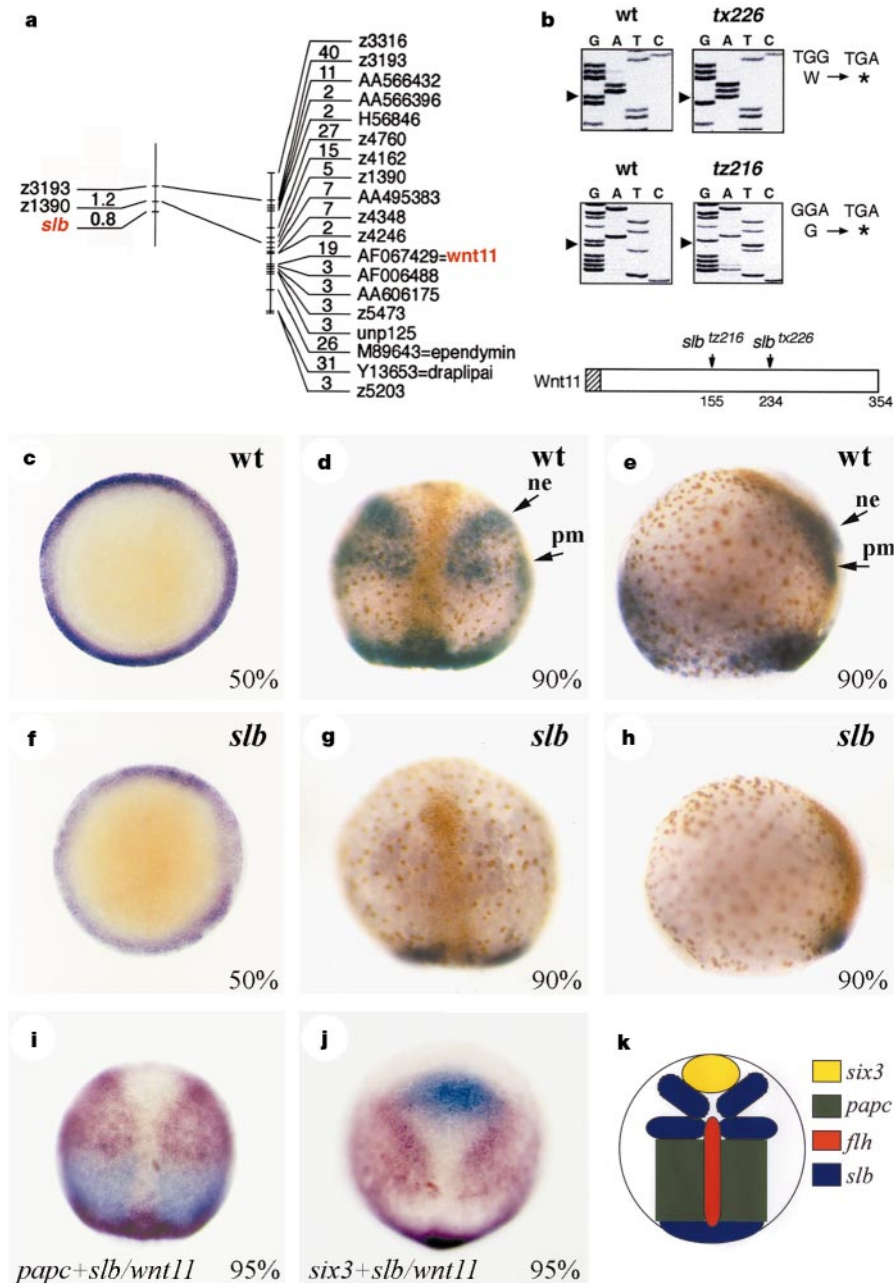


Figure 1 The *slb* locus encodes Wnt11. **a**, Positions of *slb* on a genetic map (left) and *wnt11* on a radiation hybrid map (right). **b**, Point mutations of the *slb*^{tx226} (top) and *slb*^{tz216} (bottom) alleles leading to truncations of Wnt11 (diagram at bottom). wt, wild type. **c, f**, Expression of *slb/wnt11* in wild-type (c) and *slb*^{-/-} mutant (f) embryos at 50% epiboly. Animal pole views. **d, e, g, h**, Expression of *slb/wnt11* in wild-type (d, e) and *slb*^{-/-} mutant (g, h) embryos at 90% epiboly. **d, g**, Dorsal views; **e, h**, lateral views. Fkd2

antibody stains mesendodermal cells in brown. **i, j**, Expression of *slb/wnt11* (purple) and *papc* (blue, marking the paraxial trunk mesoderm²⁹) (i) and *slb/wnt11* (purple) and *six3* (blue, marking the forebrain territory³⁰) (j) in wild-type embryos. Dorsal views. **k**, Diagram of the expression domains shown in i and j. pm, anterior paraxial mesoderm; ne, anterior lateral neuroectoderm.

is thought to block the canonical Wnt pathway^{17,18}, led to ventralization of the embryo (data not shown and ref. 17). Thus, activation and blocking of the canonical Wnt pathway leads to phenotypes not related to *slb*^{-/-}.

However, results from frogs³ indicate that, as in flies⁴, the intracellular Wnt signalling pathway might diverge downstream

of Dsh and that two domains of Dsh, the PDZ and DEP domains, might be sufficient to mediate Wnt-dependent morphogenetic events. To address which domains of Dsh are required to mediate Slb/Wnt11 activity, we injected RNA encoding an amino-terminally truncated form of Dsh (*dsh-ΔN*) that contains the PDZ and DEP domains and is thought to transduce Wnt signals that regulate

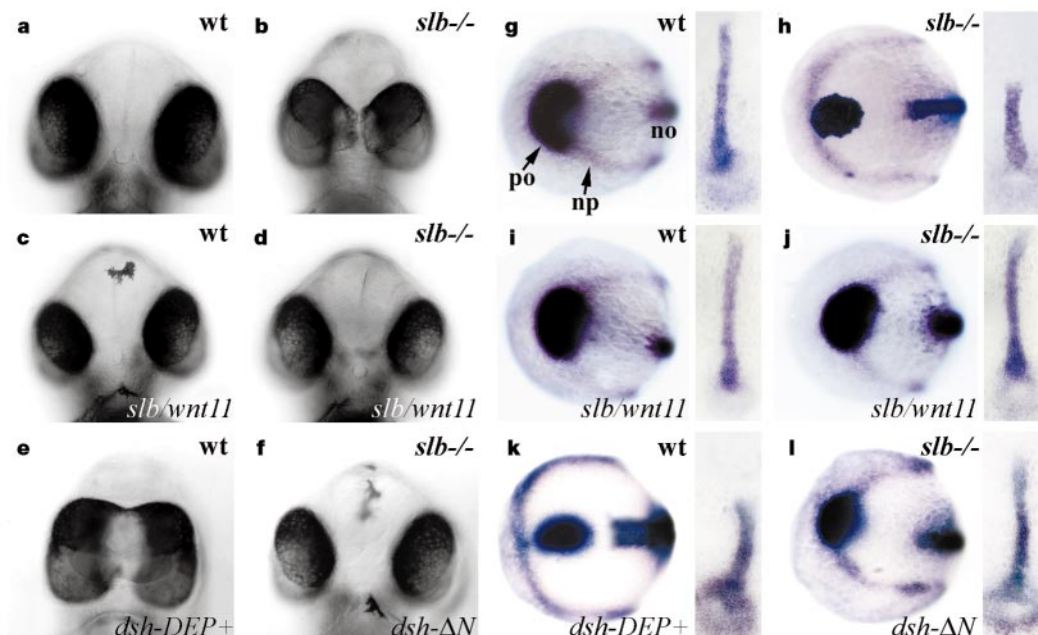


Figure 2 Injection of *slb/wnt11* and *dsh-ΔN* RNA rescues the *slb*^{-/-} mutant phenotype. Embryos were examined at pharyngula stage for spacing of the eyes (**a–f**) and at the tailbud stage for extension of the axial mesendoderm relative to the overlying anterior edge of the neural plate (**g–l**). **a–f**, Eye phenotypes at the pharyngula stage in wild-type (wt) (**a, c, e**) and *slb*^{-/-} (**b, d, f**) embryos injected with *slb/wnt11* (**c, d**), *dsh-DEP+* (**e**) and *dsh-ΔN* (**f**) RNA. Frontal views. **g–l**, Expression of *dlx3* (edge of the neural plate, np), *hgg*

(polster, po) and *ntl* (notochord, no) in wild-type (**g, i, k**) and *slb*^{-/-} (**h, j, l**) embryos injected with *slb/wnt11* (**i, j**), *dsh-DEP+* (**k**) and *dsh-ΔN* (**l**) RNA. Left of panels, animal pole views; right, notochord. *dsh-DEP+* is a dominant-interfering form of *dsh* that specifically blocks the Wnt signal pathway, affecting morphogenetic movements during *Xenopus* gastrulation, whereas *dsh-ΔN* can specifically activate this pathway³.

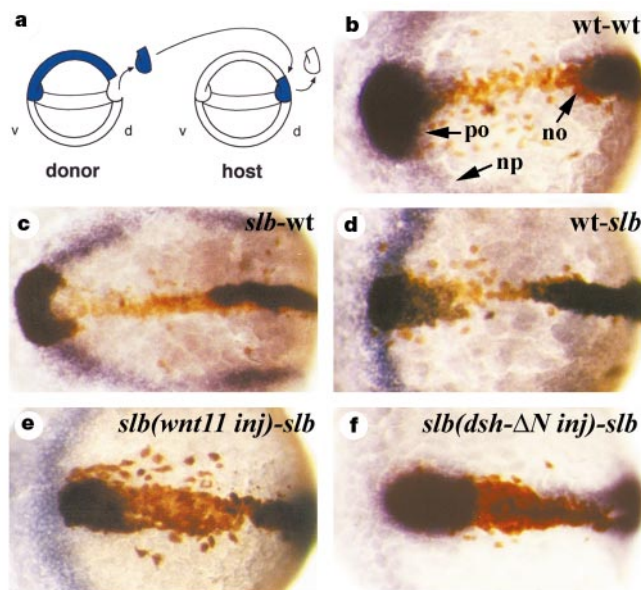


Figure 3 *slb/wnt11* is required in lateral cells for convergent extension movements during gastrulation. Embryos were examined at the tailbud stage for extension of the axial mesendoderm relative to the overlying anterior edge of the neural plate. **a**, Shield replacement procedures²¹. The shield was labelled with biotin. **b**, Wild-type (wt) shield replaced by a wild-type shield (control). **c**, Wild-type shield replaced by a *slb*^{-/-} shield.

d, *slb*^{-/-} shield replaced by a wild-type shield. **e**, *slb*^{-/-} shield replaced by a *slb*^{-/-} shield overexpressing *slb/wnt11* RNA. **f**, *slb*^{-/-} shield replaced by a *slb*^{-/-} shield overexpressing *dsh-ΔN* RNA. Tailbud-stage embryos were stained in blue for the expression of *dlx3* (edge of the neural plate, np), *hgg* (polster, po) and *ntl* (notochord, no). Transplanted cells were stained for biotin (brown). Dorsal views. d, dorsal; v, ventral.

morphogenetic movements, but not those involved in the canonical Wnt pathway³. Ubiquitous overexpression of *dsh-ΔN* RNA could rescue the *slb* phenotype, indicating that Dsh acts permissively downstream of Slb/Wnt11 (Fig. 2f, l and Tables 1 and 2). To investigate whether blocking Dsh function can phenocopy the *slb*^{-/-} phenotype, we overexpressed RNA encoding a truncated form of Dsh (Dsh-DEP+) containing the DEP domain but lacking the PDZ and DIX domains. This construct is believed to interfere with the Wnt signalling pathway, which regulates morphogenetic movements, but not the canonical Wnt signalling pathway^{3,19}. Wild-type embryos injected with *dsh-DEP+* RNA frequently showed disrupted convergent extension movements reminiscent of the phenotype of *slb* mutant embryos (Fig. 2e, k and Tables 1 and 2), further supporting the notion that the PDZ and DEP domains of Dsh mediate Slb/Wnt11 signalling.

Late aspects of the *slb*^{-/-} phenotype are believed to be a consequence of delayed and reduced anterior movement of axial midline tissue², but *slb/wnt11* is expressed primarily in non-axial tissues. This raises the possibility that Slb/Wnt11 activity is required in paraxial tissues to mediate movements of axial cells²⁰. We therefore transplanted prospective axial tissue (the embryonic shield)²¹ between wild-type and *slb*^{-/-} embryos and tested whether wild-type Slb/Wnt11 activity was required in axial or paraxial cells for normal axial cell movements (Fig. 3a, b; results summarized in Table 3). The behaviour of *slb*^{-/-} shield derivatives in an otherwise wild-type embryo was indistinguishable from that of wild-type shield

derivatives (Fig. 3c). Conversely, wild-type shield derivatives in an otherwise *slb* mutant embryo behaved like *slb*^{-/-} shield derivatives in a *slb*^{-/-} embryo (Fig. 3d). These results indicate that the expression of *slb* in paraxial cells might be required for normal convergent extension of midline tissues.

To examine whether Slb/Wnt11 secreted from lateral cells might act directly on midline cells to mediate cell movement, we replaced the shields of *slb*^{-/-} embryos with shields from other *slb*^{-/-} embryos that had been injected with RNA encoding *slb/wnt11* or *dsh-ΔN* at the one-cell stage. Neither the ectopic expression of *slb/wnt11* nor that of *dsh-ΔN* in *slb*^{-/-} shield derivatives restored wild-type convergent extension movements of axial tissue (Table 3 and Fig. 3e, f). These results indicate that Slb/Wnt11 activity in midline cells alone might not be sufficient for normal convergent extension movements of axial tissue.

To address how cell movement is disrupted during gastrulation in *slb*^{-/-} embryos, we transplanted small groups of lineage-labelled non-axial mesodermal cells into *slb*^{-/-} and wild-type embryos (Fig. 4a) and assessed the movements and final location of the transplanted cells. Throughout these experiments no difference was observed between grafts of wild-type and *slb*^{-/-} cells. In wild-type embryos (*n* ≥ 35), wild-type or *slb*^{-/-} cells transplanted adjacent to the shield formed strings of cells extended along the antero-posterior axis adjacent to anterior axial mesoderm (Fig. 4d, top embryo). Cells transplanted to more lateral positions gave rise to strings of cells located in anterior or posterior non-axial mesoderm,

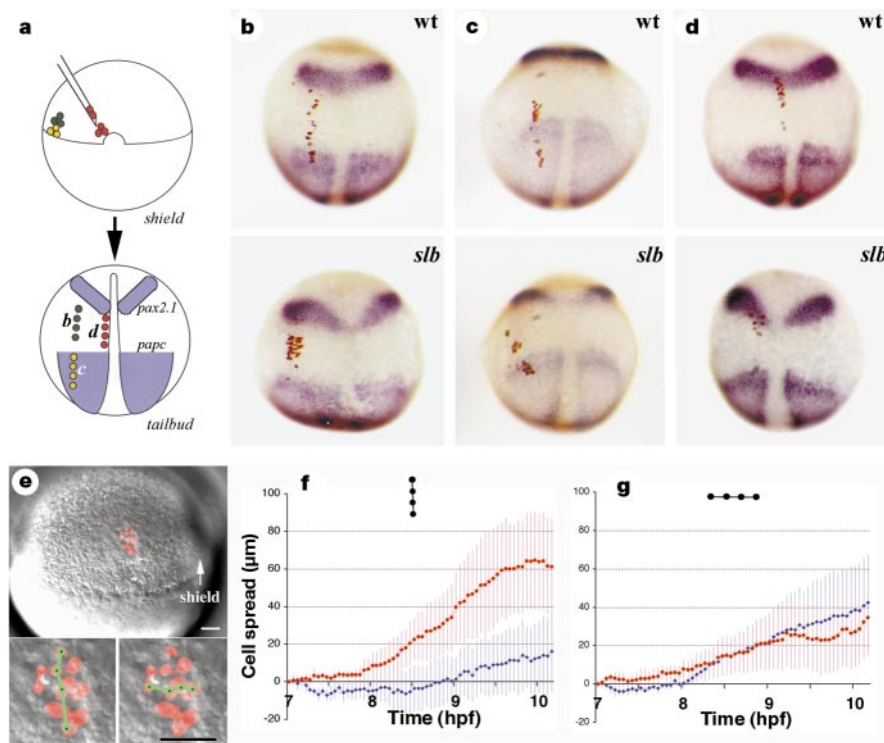


Figure 4 Convergent extension movements of paraxial mesodermal cells are affected in *slb*^{-/-} mutant embryos. **a**, Cell transplantation procedures. Cells were transplanted deep into the margin either close to the shield (red cells in **a**) or 60° lateral to the shield both directly at the margin (yellow cells in **a**) or slightly anterior to it (green cells in **a**) at the shield stage (6 h after fertilization). The positions of the transplanted cells were determined at the tailbud stage (10 h after fertilization) relative to the expression domains of *pax2.1*, which lies anterior of the paraxial mesodermal expression domain of *slb/wnt11*, and *papc*, which defines the posterior boundary (see also Fig. 1k). **b**, Embryos in which cells (prospective anterior paraxial mesoderm) were transplanted into the margin ~60° lateral to the shield (top, wild-type (wt) cells into a wild-type embryo; bottom, *slb*^{-/-} cells in a *slb*^{-/-} embryo). **c**, Embryos in which cells (prospective posterior paraxial mesoderm) were transplanted into the margin 60° lateral to the shield (top, wild-type cells into a wild-type

embryo; bottom, *slb*^{-/-} cells in a *slb*^{-/-} embryo). **d**, Embryos in which cells (prospective anterior paraxial mesoderm) were transplanted into the blastula margin close to the shield (top, wild-type cells into a wild-type embryo; bottom, *slb*^{-/-} cells in a *slb*^{-/-} embryo). Transplanted cells are labelled with biotin (brown). **e**, Lateral view of a living embryo 7 h after fertilization (top, animal pole) revealed with Nomarski optics, showing a group of transplanted rhodamine-labelled cells (in red) within the anterior paraxial mesoderm. **f**, **g**, A group of four cells (see insets in **e**) arranged in either columns (**f**) or rows (**g**) was tracked over time (hours after fertilization, hpf) to determine the degree of cell spread along the antero-posterior and medio-lateral axes, respectively. For each group, the mean cell distance (and s.d.) between pairs of cells was plotted in wild-type (red) and *slb*^{-/-} (blue) embryos. Scale bars, 50 μm.

Table 3 *slb/wnt11* is required in lateral cells to regulate convergent extension

Donor	Host	Wild type (%)	<i>slb</i> (%)	Total (n)
wild type	wild type	100	0	9
wild type	<i>slb</i>	4	96	18
<i>slb</i>	wild type	93	7	30
<i>slb (wnt11)</i>	<i>slb</i>	2	98	41
<i>slb(dsh-ΔN)</i>	<i>slb</i>	0	100	20

Normal convergent extension movements were classified as either 'wild type' or '*slb*' by the shape and position of the prechordal plate in relation to the anterior edge of the neural plate. The notochord often showed variable abnormalities after shield transplantations and was therefore not used to classify the embryos.

depending on the initial position of the transplant (Fig. 4b, c, top embryos). When comparable transplants of wild-type or *slb*^{-/-} cells were performed in *slb*^{-/-} embryos ($n \geq 35$), cells ended up in the correct regions of the mesoderm, but the clones failed to form the elongated strings observed in wild-type hosts (Fig. 4b–d, bottom embryos). These results suggest that the medio-lateral cell intercalations associated with convergent extension are disrupted in *slb*^{-/-} embryos. Furthermore, the observation that the genotype of the transplanted cells did not affect their final position indicates that Slb/Wnt11 functions in a cell non-autonomous manner throughout the non-axial mesoderm.

To assess directly whether cell intercalation is disrupted in *slb*^{-/-} embryos, we repeated the experiments described above and measured the movements and behaviour of mesodermal cells in living wild-type and *slb*^{-/-} embryos (Fig. 4e). In wild-type embryos, mesodermal cells spread rapidly along the antero-posterior axis during the last 2 h of gastrulation (8–10 h after fertilization) (Fig. 4f, red). At this time, many unlabelled mesodermal cells intercalate between the initially coherent clone of labelled cells, increasing the distance between these cells. In *slb*^{-/-} embryos, the rapid dispersal of cells along the antero-posterior axis was much less apparent (Fig. 4f, blue), whereas cell movements towards the dorsal side of the embryo resembled the wild-type condition (data not shown). There was also no significant difference between wild-type and *slb*^{-/-} embryos in the distribution of cells relative to each other along the medio-lateral axis (Fig. 4g). These observations indicate that Slb/Wnt11 is required in non-axial mesoderm to mediate cell intercalation along the antero-posterior axis that contributes to the extension of the body axis during late gastrulation.

Our results provide experimental and genetic evidence that *Wnt* genes regulate convergent extension movements in the nascent non-axial mesoderm during gastrulation. Our finding that Dsh acts downstream of Slb/Wnt11 via a non-canonical Wnt pathway reveals similarities between the intracellular pathways regulating convergent extension movements in vertebrates and planar polarity in *Drosophila*^{19,22}. Indeed, it raises the possibility that Wnt-mediated assignment of cell polarity might contribute to the morphogenetic movements of gastrulation. Furthermore, the expression of *slb/wnt11* in non-axial ectoderm, coupled with the observations that comparable extension movements occur in the neural plate^{23,24}, indicates a possible global role for Wnt mediated regulation of cell behaviour in all germ layers.

Further analysis of other zebrafish mutants showing defects in convergent extension movements might help in the elucidation of the Wnt signalling pathway regulating morphogenesis. The observation that *slb* expression is unchanged in *trilobite* (*tri*) mutant embryos^{7,8} (M.T. and C.P.H., unpublished observations) raises the possibility that *Tri* is a downstream mediator of Slb/Wnt11 function. □

Methods

Fish lines and genetics

Two alleles of *slb* (*tx226* and *tz216*) were used for sequence analysis. All injection and transplantation experiments were performed with embryos from homozygous *slb*^{tx226} carriers.

Linkage analysis, mapping and cloning

slb^{tz216} was mapped in F₂ offspring of a Tü × WIK reference cross as described²⁵, with the use of Simple Sequence Length Polymorphism markers²⁶ on pools of 48 mutants and 48 siblings. Linkages from the pools were confirmed and refined by genotyping single embryos. To identify mutations, RNA from shield-stage embryos (6 h after fertilization) of crosses between *slb*^{tx226}, *slb*^{tz216} homozygous carriers or wild-type fish was reverse-transcribed, then amplified by PCR with *pfu* polymerase for sequencing analysis. The mutations were confirmed on genomic DNA of both alleles.

RNA injections and constructs

For RNA injection, *slb* constructs were generated by polymerase chain reaction (PCR) with complementary DNA from wild-type or *slb*^{tx226} embryos. The PCR fragments were cloned into pCS2+. The *dsh* constructs *dsh-ΔN* and *dsh-DEP+* were used as described³. RNA was injected into one-cell-stage embryos as described²⁷.

Whole-mount antibody labelling and *in situ* hybridization

Whole-mount *in situ* hybridization was performed essentially as described²⁷. Double *in situ* hybridization was done with fluorescein- and digoxigenin-labelled probes and detected with 5-bromo-4-chloro-3-indolylphosphate alone for the first probe and BM purple for the second probe.

Tissue and cell transplantation

For transplantations, donor embryos were injected with a mixture of fluorescent and biotinylated dextrans, and cells transplanted into host embryos were detected with a Vectastain kit to reveal biotin¹. For shield transplantations, the shield was removed from a shield-stage donor embryo with an aspiration pipette connected to a microsyringe pump, and inserted under the enveloping layer of a host embryo at the same stage²¹.

Time-lapse imaging and cell tracking

For time-lapse imaging, fluorescence-labelled cells were transplanted as above. Embryo preparation and image acquisition were performed as described²⁸ but with OPENLAB software, a 20× water immersion objective and a Zeiss Axioplan microscope. The movement of groups of cells arranged along either the antero-posterior (columns) or the medio-lateral (rows) axes was tracked every 3 min and the distance between pairs of cells within each group was measured with customized routines written for NIH Image²⁸. For each time point, the mean cell distance and s.d. were determined, normalized to the initial measurement (for example, distance at 7 h after fertilization) and plotted in Excel.

Received 17 January; accepted 18 February 2000.

- Heisenberg, C. P. et al. Genes involved in forebrain development in the zebrafish, *Danio rerio*. *Development* **123**, 191–203 (1996).
- Heisenberg, C. P. & Nüsslein-Volhard, C. The function of *silberblick* in the positioning of the eye anlage in the zebrafish embryo. *Dev. Biol.* **184**, 85–94 (1997).
- Tada, M. & Smith, J. C. *Xwnt11*, a target of *Xenopus* Brachyury: Regulation of gastrulation movements via Dishevelled, but not through the canonical Wnt pathway. *Development* (in the press).
- Boutros, M. & Mlodzik, M. Dishevelled: at the crossroads of divergent intracellular signaling pathways. *Mech. Dev.* **83**, 27–37 (1999).
- Warga, R. M. & Kimmel, C. B. Cell movements during epiboly and gastrulation in zebrafish. *Development* **108**, 569–580 (1990).
- Keller, R., Shih, J. & Domingo, C. The patterning and functioning of protrusive activity during convergence and extension of the *Xenopus* organizer. *Development* (Suppl.), 81–91 (1992).
- Hammerschmidt, M. et al. Mutations affecting morphogenesis during gastrulation and tail formation in the zebrafish, *Danio rerio*. *Development* **123**, 143–151 (1996).
- Solnica-Krezel, L. et al. Mutations affecting cell fates and cellular rearrangements during gastrulation in zebrafish. *Development* **123**, 67–80 (1996).
- Moon, R. T. et al. *Xwnt-5A*: a maternal Wnt that affects morphogenetic movements after over-expression in embryos of *Xenopus laevis*. *Development* **119**, 97–111 (1993).
- Du, S. J., Purcell, S. M., Christian, J. L., McGrew, L. L. & Moon, R. T. Identification of distinct classes and functional domains of Wnts through expression of wild-type and chimeric proteins in *Xenopus* embryos. *Mol. Cell. Biol.* **15**, 2625–2634 (1995).
- Ungar, A. R., Kelly, G. M. & Moon, R. T. Wnt4 affects morphogenesis when misexpressed in the zebrafish embryo. *Mech. Dev.* **52**, 153–164 (1995).
- Makita, R., Mizuno, T., Koshida, S., Kuroiwa, A. & Takeda, H. Zebrafish Wnt11—pattern and regulation of the expression by the yolk cell and no tail activity. *Mech. Dev.* **71**, 165–176 (1998).
- Cadigan, K. M. & Nusse, R. Wnt signaling—a common theme in animal development. *Genes Dev.* **11**, 3286–3305 (1997).
- Moon, R. T. & Kimelman, D. From cortical rotation to organizer gene expression—toward a molecular explanation of axis specification in *Xenopus*. *BioEssays* **20**, 536–545 (1998).
- Glinka, A., Wu, W., Onichtchouk, D., Blumenstock, C. & Niehrs, C. Head induction by simultaneous repression of Bmp and Wnt signalling in *Xenopus*. *Nature* **389**, 517–519 (1997).
- Pai, L. M., Orsulic, S., Bejsovec, A. & Peifer, M. Negative regulation of Armadillo, a wingless effector in *Drosophila*. *Development* **124**, 2255–2266 (1997).
- Plegri, F. & Maischein, H. M. Function of zebrafish beta-catenin and Tcf-3 in dorsoventral patterning. *Mech. Dev.* **77**, 63–74 (1998).
- Molenaar, M. et al. Xtc-3 transcription factor mediates beta-catenin-induced axis formation in *Xenopus* embryos. *Cell* **86**, 391–399 (1996).
- Axelrod, J. D., Miller, J. R., Shulman, J. M., Moon, R. T. & Perrimon, N. Differential recruitment of dishevelled provides signaling specificity in the planar cell polarity and Wingless signaling pathways. *Genes Dev.* **12**, 2610–2622 (1998).
- Wilson, P. A., Oster, G. & Keller, R. Cell rearrangement and segmentation in *Xenopus*: direct

- observation of cultured explants. *Development* **105**, 155–166 (1989).
21. Saude, L., Woolley, K., Martin, P., Driever, W. & Stemple, D. L. Axis inducing activity and cell fates of the zebrafish organizer. *Development* (submitted).
 22. Boutros, M., Paricio, N., Strutt, D. I. & Mlodzik, M. Dishevelled activates JNK and discriminates between JNK pathways in planar polarity and wingless signaling. *Cell* **94**, 109–118 (1998).
 23. Kimmel, C. B., Warga, R. M. & Kane, D. A. Cell cycles and clonal strings during formation of the zebrafish central nervous system. *Development* **120**, 265–276 (1994).
 24. Elul, T., Koehl, M. A. & Keller, R. E. Cellular mechanisms underlying neural convergent extension in *Xenopus laevis* embryos. *Dev. Biol.* **191**, 243–258 (1997).
 25. Rauch, G.-J., Granato, M. & Haffter, P. A polymorphic zebrafish line for genetic mapping using SSLPs on high-percentage agarose gels. *Tech. Tips Online* T01208 (cited 080300) (<http://www.biomednet.com/db/tto>) (1997).
 26. Knapik, E. W. *et al.* A reference cross DNA panel for zebrafish (*Danio rerio*) anchored with simple sequence length polymorphisms. *Development* **123**, 451–460 (1996).
 27. Barth, K. A. & Wilson, S. W. Expression of zebrafish *nkx2.2* is influenced by sonic hedgehog/vertebrate hedgehog-1 and demarcates a zone of neuronal differentiation in the embryonic forebrain. *Development* **121**, 1755–1768 (1995).
 28. Concha, M. L. & Adams, R. J. Oriented cell divisions and cellular morphogenesis in the zebrafish gastrula and neurula: a time-lapse analysis. *Development* **125**, 983–994 (1998).
 29. Yamamoto, A. *et al.* Zebrafish paraxial protocadherin is a downstream target of *spadetail* involved in morphogenesis of gastrula mesoderm. *Development* **125**, 3389–3397 (1998).
 30. Kobayashi, M., Toyama, R., Takeda, H., Dawid, I. B. & Kawakami, K. Overexpression of the forebrain-specific homeobox gene *six3* induces rostral forebrain enlargement in zebrafish. *Development* **125**, 2973–2982 (1998).

Acknowledgements

We thank T. Schilling, C. Houart, P. Haramis and F. Conlon for critical reading of this manuscript, many colleagues for providing reagents, and the late P. Haffter for providing data before publication. C.P.H. and M.L.C. were supported by postdoctoral fellowships from EMBO, the EC and the Wellcome Trust. G.J.R. and R.E.G. were supported by a grant from the German Human Genome Project. L.S. was supported by a PhD. grant from the Fundação para a ciência e a Tecnologia, Programa Praxis XXI. J.C.S. and D.L.S. were supported by the MRC. This work was also supported by a Human Frontier Science Programme grant to J.C.S., S.W.W., H. Sive and N. Ueno. S.W.W. is a Wellcome Trust Senior Research Fellow supported by the BBSRC and the Wellcome Trust.

Correspondence and requests for materials should be addressed to C.P.H. (e-mail: c.heisenberg@ucl.ac.uk) or M.T. (e-mail: mtada@nimr.mrc.ac.uk).

Dishevelled controls cell polarity during *Xenopus* gastrulation

John B. Wallingford*, Brian A. Rowning†, Kevin M. Vogeli*, Ute Rothbacher‡§, Scott E. Fraser ‡ & Richard M. Harland*

* Department of Molecular and Cell Biology, 401 Barker Hall, University of California, Berkeley, California 94720-3204, USA

† Life Sciences Division (MS6-2100), Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

‡ Division of Biology and Beckman Institute, California Institute of Technology, Pasadena, California 91125, USA

Although cell movements are vital for establishing the normal architecture of embryos, it is unclear how these movements are regulated during development in vertebrates. Inhibition of *Xenopus* Dishevelled (Xdsh) function disrupts convergent extension movements of cells during gastrulation, but the mechanism of this effect is unclear, as cell fates are not affected¹. In *Drosophila*, Dishevelled controls both cell fate and cell polarity^{2–4}, but whether Dishevelled is involved in controlling cell polarity in vertebrate embryos has not been investigated. Here we show, using time-lapse confocal microscopy, that the failure of cells lacking Xdsh function to undergo convergent extension results from defects in cell polarity. Furthermore, Xdsh mutations that inhibit convergent extension correspond to mutations in *Drosophila* Dishevelled that selectively perturb planar cell polarity. Finally, the localization of Xdsh at the membrane of normal dorsal mesodermal cells is consistent with Xdsh controlling cell polarity. Our results show that polarized cell behaviour is essen-

tial for convergent extension and is controlled by vertebrate Dishevelled. Thus, a vertebrate equivalent of the *Drosophila* planar cell polarity signalling cascade may be required for normal gastrulation.

The morphogenetic cell movements of gastrulation are critical for establishing the germ layers and embryonic axes in animals, and convergent extension movements are important during gastrulation^{5,6}. During *Xenopus* convergent extension, lamellipodia that extend from dorsal mesodermal cells become mediolaterally polarized and make stable contacts with neighbouring cells. Alignment, elongation and intercalation of cells follows, leading to narrowing of the mediolateral axis (convergence) and lengthening of the anteroposterior axis (extension) (Fig. 1b)^{7–9}. The molecular mechanisms controlling these polarized cell behaviours are unclear.

Dishevelled (Dsh) and other members of the wingless (Wg)/Wnt pathway have many roles in embryogenesis¹⁰, and dorsal translocation of Xdsh in *Xenopus* is an essential event in establishing dorsal cell fates¹¹. Expression of a mutant form of Xdsh, Xdd1, inhibits convergent extension of the dorsal mesoderm, but these defects are specific to morphogenesis and the embryos display no defects in dorsoventral axis specification¹. These results indicate that Xdsh may have two distinct functions in vertebrate embryogenesis, consistent with the situation in *Drosophila*, where Dsh signalling modulates cell fate through the Wg/Wnt pathway and also mediates cell polarity through the planar cell polarity (PCP) cascade^{2–4} (Fig. 1d). However, Dsh has not previously been shown to mediate cell polarity decisions in any vertebrate embryo⁴.

We were interested in how polarized cell behaviour in the *Xenopus* dorsal marginal zone (DMZ) was affected by modulation of

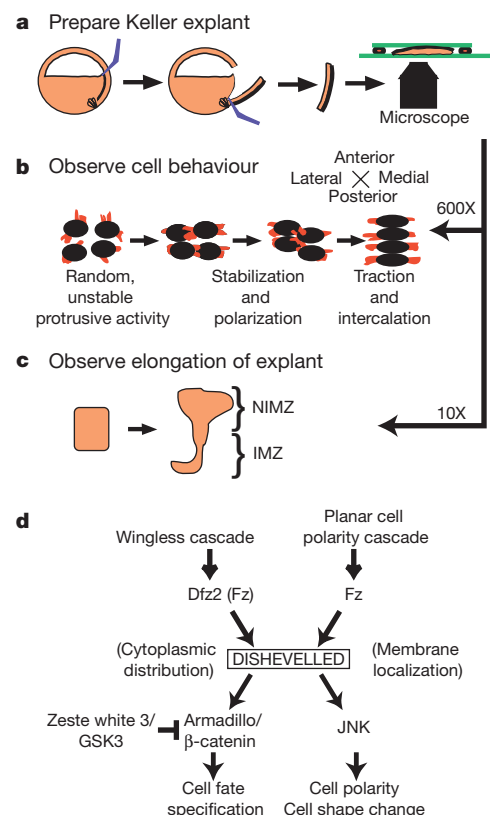


Figure 1 Methods and signalling pathways. **a**, Preparation of Keller explants. DMZs are removed at the onset of gastrulation and immobilized in culture under coverslips. **b**, Time-lapse observation at high magnification allows examination of cell behaviour (black, cell bodies; red, lamellipodia). **c**, Observation at low magnification allows assessment of elongation and morphogenesis; the neural ectoderm (NIMZ) does not elongate; the mesoderm (IMZ) does⁷. **d**, Dishevelled mediates both Wg and PCP signalling cascades⁴.

§ Present address: Laboratoire de Genetique et Physiologie du Developpement, IBDM, Campus de Luminy, Marseille, France

Dishevelled signalling. Overexpression of Xdd1 inhibits convergent extension of the DMZ¹. Likewise, overexpression of wild-type Xdsh significantly inhibits convergent extension (see below), consistent with results from *Drosophila*, where overexpression of wild-type Dishevelled elicits dominant disruption of planar cell polarity. We used confocal time-lapse microscopy of Keller explants (Fig. 1a, b) to examine the cell behaviour underlying Xdd1- and Xdsh-mediated disruption of convergent extension.

As expected⁹, analysis of control explants expressing a membrane-targeted green fluorescent protein (GFP) confirmed that most lamellipodia in control DMZ explants were stable (Fig. 2a–e, l). In contrast, dorsal mesodermal cells expressing Xdd1 extended and withdrew significantly more protrusions and maintained significantly fewer stable protrusions than did controls (Fig. 2f–l; time-lapse video images available at <http://mcb.berkeley.edu/labs/harland/>). Thus, cells expressing Xdd1 are not defective in their ability to make membrane protrusions, but are unable to maintain them stably. On the other hand, cells overexpressing Xdsh had no defects in lamellipodial stability (Fig. 2l), indicating that other defects in cell behaviour must underlie the inhibition of convergent extension.

We next examined cell polarity in Keller explants by determining the orientation of membrane protrusions from individual cells.

Whereas lamellipodia were predominantly extended from the mediolateral ends of control cells, this polarity was severely disrupted in both Xdd1- and Xdsh-expressing cells (Fig. 2m).

Analysis of cell elongation and orientation of the long axis provide additional indicators of cell polarity in the DMZ^{7,8}. As cells begin convergent extension, they align and elongate in the mediolateral axis (Fig. 2n); the mean length-to-width ratio (LWR) of control cells was $1.97 (\pm 0.31)$, similar to that reported⁸. On the other hand, cells expressing Xdd1 (Fig. 2o; LWR = 1.65 ± 0.12) or Xdsh (Fig. 2p; LWR = 1.57 ± 0.08) were not elongated. These LWRs are similar to those of DMZ cells before gastrulation or animal cap cells, which do not converge and extend⁸. Thus, cell elongation is correlated with normal, but not inhibited, convergent extension.

In Fig. 2q, the LWR of each cell is plotted against the angle of its long axis. This plot illustrates the bias of elongate control cells (red diamonds) towards the mediolateral axis and the random angular distribution of cells expressing Xdd1 (black squares) or Xdsh (blue circles). Over 67% of control cells have their long axes in the mediolateral sector (Fig. 2n, q, r). This polarity is severely disrupted in Xdd1-expressing explants (Fig. 2o, q, r) and in Xdsh-expressing explants (Fig. 2p–r). These data show that cell polarity is severely disrupted in Xdd1- and Xdsh-expressing DMZ explants.

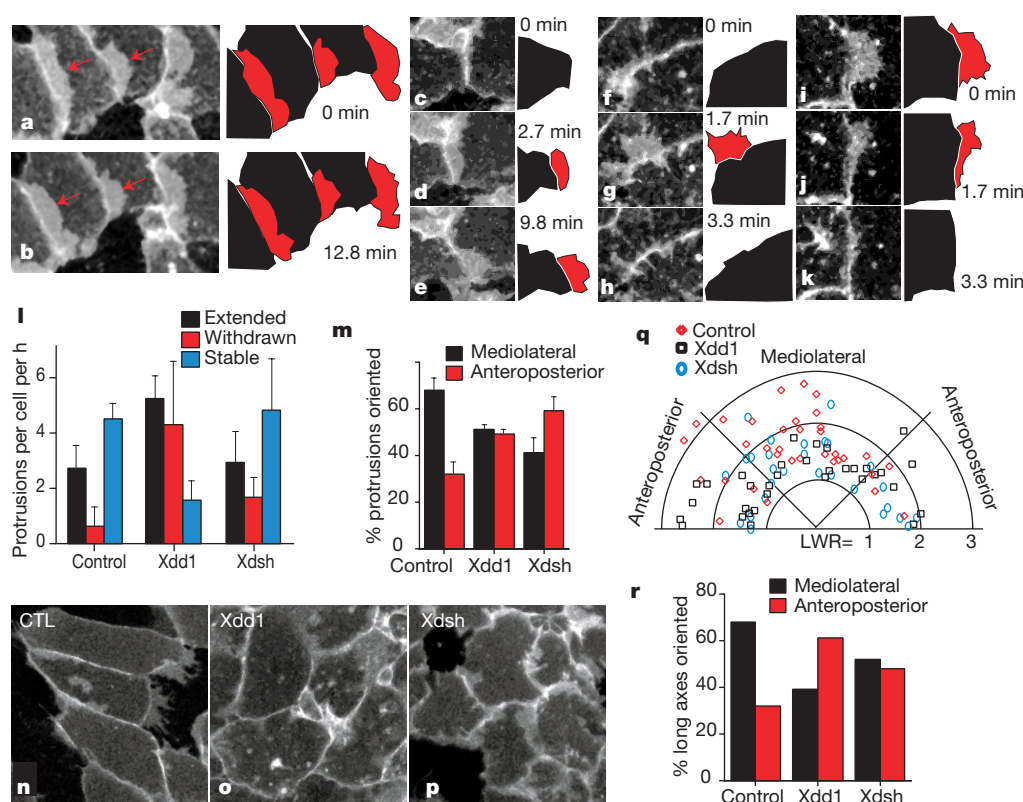


Figure 2 Analysis of cell behaviour and polarity in Keller explants. Overexpression of Xdd1 or Xdsh severely disrupts cell polarity. **a–j**, Left, single time-lapse frames show representative cell behaviours; right, diagrams of cells with time points. Cell bodies, black; lamellipodia, red; horizontal is mediolateral. **a, b**, Arrows, stable mediolaterally biased lamellipodia in control DMZ cells. **c–e**, Lamellipodium extending and stabilizing mediolaterally in a control explant. **f–h**, Xdd1-expressing cell quickly extending and retracting a posteriorly biased lamellipodium. **i–k**, Xdd1-expressing cell quickly retracting a mediolaterally biased lamellipodium. **l**, Protrusive activity: control cells make and withdraw few protrusions but maintain many stable ones. Xdd1-expressing cells make and retract significantly more protrusions than do controls and maintain fewer stable protrusions. Xdsh-expressing cells are stable. **m**, Orientation of protrusions. Lamellipodia

are mediolaterally biased in control cells but randomly distributed in cells expressing Xdd1 or Xdsh. Values shown in **l** and **m** are means $\pm$ s.d. between explants. Differences are statistically significant ($P < 0.01$) by the Mann–Whitney *U*-test (control, 33 cells/4 explants; Xdd1, 32 cells/4 explants; Xdsh, 29 cells/3 explants). **n**, Mediolaterally aligned, elongating control cells. Xdd1- (**o**) and Xdsh-expressing cells (**p**) fail to align or elongate. **q**, LWR versus angle of major axis for each of the cells scored in **l**, **m**. Elongate control cells are clustered mediolaterally; Xdd1- and Xdsh-expressing cells are randomly distributed. **r**, Per cent of cells oriented in mediolateral and anteroposterior sectors (derived from **q**). Control cells are mediolaterally biased; Xdsh- and Xdd1-expressing cells are not.

Although polarized cell behaviour has been shown to be associated with convergent extension, our results show that experimental disruption of cell polarity is directly associated with failure of both cell elongation and convergent extension. In light of the observed defects, we investigated the possibility that Xdsh may control cell polarity in a manner similar to that of Dsh in *Drosophila*.

Drawing on the separable functions of Dsh²⁻⁴ (Fig. 1d), we tested reagents that differentially modulate signalling in the Wnt and PCP pathways for their ability to affect convergent extension. We assessed convergent extension by observing changes in both length and shape of isolated Keller explants¹² (Fig. 1c). Control explants elongated significantly and underwent typical changes in morphology between the gastrula and neurula stages (Fig. 3a, b), and expression of Xdd1 inhibited both elongation and morphological changes of explants (Fig. 3a, c). This effect was partially prevented by co-expression of wild-type Xdsh (Fig. 3a, d), indicating that the effect was specific.

To investigate whether canonical Wnt signalling is required during convergent extension, we tested whether specific activation of this pathway downstream of Dsh could rescue the effects of Xdd1. Dominant-negative (DN)-GSK3 is a potent downstream activator of the Wnt pathway¹³ (Fig. 1d), but co-expression of DN-GSK3 with Xdd1 did not rescue convergent extension (Fig. 3a, e), indicating that downstream activation of the canonical Wnt pathway does not reverse the effects of Xdd1 on convergent extension. Similar results were obtained in intact embryos (J.B.W. and R.M.H., unpublished data). Overexpression of DN-GSK3 alone in the DMZ did not significantly affect convergent extension (Fig. 3a), but overexpression of wild-type Xdsh alone strongly inhibited convergent extension (Fig. 3a, f). In *Drosophila*, overexpression of wild-type Dsh

produces a planar polarity defect, but does not affect Wg/Wnt signalling². Thus the contrasting effects of DN-GSK3 and Xdsh (both of which activate the Wnt pathway) indicate that the control of cell polarity in the DMZ by Xdsh is largely independent of canonical Wg/Wnt signalling.

To investigate whether Xdsh signals through a PCP-like cascade in the *Xenopus* DMZ, we used deletions of Xdsh similar to those that uncouple Wingless and planar polarity phenotypes in *Drosophila*²⁻⁴. Xdsh-ΔPDZ (a PDZ domain deletion similar, but not identical, to Xdd1) and Xdsh-ΔDEP (a deletion of the DEP domain) are both functional for Wnt signalling in *Xenopus* and *Drosophila*, but have dominant, inhibitory effects on planar polarity in *Drosophila*^{2,14} (Fig. 4). Expression of either Xdsh-ΔPDZ or Xdsh-ΔDEP strongly inhibited convergent extension in Keller explants (Fig. 3a, g, h). These phenotypes were partially rescued by co-expression of wild-type Xdsh (Fig. 3a). As both Xdsh-ΔPDZ and Xdsh-ΔDEP can activate the Wnt pathway (Fig. 4), it is important to note that the induced failure of convergent extension was not the result of excessive Wnt signalling; co-expression of additional wild-type Xdsh (which is fully active for both Wnt and PCP signalling) did not exacerbate the convergent extension phenotypes of the mutant Xdsh constructs, but instead ameliorated these effects (Fig. 3a).

Another deletion of Dsh, Xdsh-ΔDIX, has only very weak effects on PCP signalling in *Drosophila* assays (Fig. 4), and likewise only very weakly affects convergent extension (Fig. 3a); most explants expressing Xdsh-ΔDIX do change shape and elongate (Fig. 3i). This result is particularly telling, as Xdsh-ΔDIX and Xdd1 have very different effects on convergent extension but both have equivalent weak ability to activate the Wnt signalling pathway in a *Xenopus*

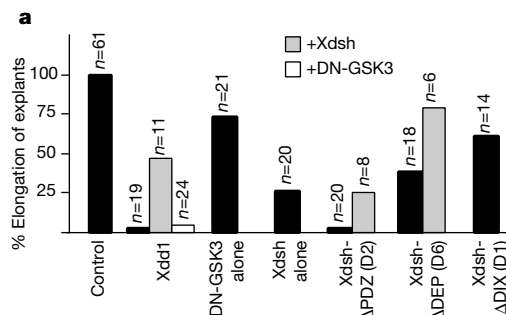
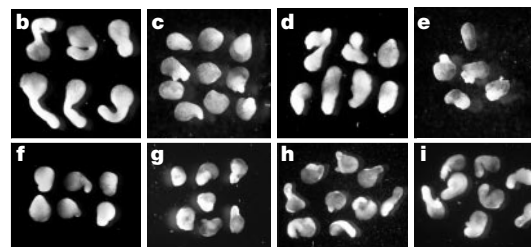


Figure 3 Analysis of convergent extension in Keller explants. **a**, Relative elongation of DMZ explants compared with controls. Constructs expressed are indicated below graph; co-injected constructs are indicated in legend above graph. All differences shown are statistically significant ($P < 0.01$) by the Mann-Whitney *U*-test. **c–i**, Keller explants from representative experiments. **b**, DMZ explants from uninjected embryos elongate substantially between stages 10+ and 19, reorganizing into an IMZ 'tail' and a NIMZ 'head' (Fig. 1c). **c**, Explants expressing Xdd1 fail to elongate. Co-expression of wild-type



Xdsh (**d**) partially rescues the phenotype; co-expression of DN-GSK3 does not (**e**); ventral expression of DN-GSK3 induces ectopic axes (not shown), indicating that the dose used activated Wnt signalling. Explants overexpressing wild-type Xdsh (**f**), Xdsh-ΔPDZ (**g**) or Xdsh-ΔDEP (**h**) fail to elongate. Effects of mutant Xdsh constructs were rescued by co-expression with wild-type Xdsh (**a**). Most explants expressing Xdsh-ΔDIX (**i**) change shape normally and extend the IMZ: overall elongation is only slightly reduced.

	DIX	PDZ	DEP	<i>Xenopus</i> 2° axis	<i>Drosophila</i> Wg assay	<i>Drosophila</i> PCP assay	<i>Xenopus</i> convergent extension
Xdsh*	++	++	++	++	++	++	++
Xdsh-ΔPDZ	++	++	++	++	++	++	++
Xdsh-ΔDEP	++	++	++	++	++	++	++
Xdsh-ΔDIX	++	++	++	++	++	++	++
Xdd1	++	++	++	++	nt	nt	++

Figure 4 Xdsh constructs and their effects in Wnt, PCP and convergent extension assays. ++, strongly active; +, weakly active; -, weakly inhibitory; —, strongly inhibitory; nt, not

tested. *Xenopus* secondary axis assay from ref. 14; *Drosophila* assays from refs 2, 3; convergent extension assay from Fig. 3a. Asterisk, Dsh hyperactivation phenotype.

secondary axis assay (Figs 3a, 4), indicating again that inhibition of convergent extension is not a result of activating Wnt signalling.

As *Xdsh* regulates cell polarity in the DMZ (Fig. 2) and its effects are independent of Wnt signalling (Fig. 3), there should be evidence of *Xdsh* signalling through the PCP pathway in normal DMZ cells. To identify such activity, we made use of the fact that activation of the PCP pathway localizes Dsh to the inner surface of the cell membrane, whereas activation of the canonical Wnt pathway leaves Dsh in the cytoplasm². In posterior dorsal mesoderm cells that have not yet begun to converge and extend⁸, an *Xdsh*–GFP construct was localized to the cytoplasm (Fig. 5a). A similar distribution was observed in ventral marginal zone cells (not shown), which do not converge and extend. On the other hand, in anterior chordamesoderm cells that were actively engaged in convergent extension, the *Xdsh*–GFP signal was localized to the cell membrane (Fig. 5b). Furthermore, *Xdsh*–GFP was found in the cytoplasm in naive animal cap cells, which do not converge and extend (Fig. 5c). But in animal cap cells treated with activin to induce convergent extension¹⁵, *Xdsh*–GFP was present at cell membranes (Fig. 5d). This localization of *Xdsh* to the cell membrane during convergent extension is consistent with *Xdsh* regulating cell polarity through a PCP-like cascade.

Although the role of Dsh in mediating Wnt signalling and cell fate has been extensively examined, its ability to mediate cell polarity during vertebrate development has remained largely unexplored. Vertebrate Dsh signals through a cascade similar to the *Drosophila*

PCP pathway¹⁶, but no *in vivo* function for this activity has been identified in any vertebrate. Here we show that Dishevelled controls essential cell polarity decisions during convergent extension in *Xenopus*. Furthermore, analysis of *Xdsh* deletion mutants and *Xdsh* subcellular localization indicate that *Xdsh* signals through a vertebrate PCP cascade, indicating a common mechanism for vertebrate and *Drosophila* Dishevelled in regulating cell polarity. Coupled with its important role in early specification of the dorsoventral axis¹¹, these results place vertebrate Dishevelled at the nexus of the two events critical to establishing the normal architecture of the embryo, specification of cell fate and coordination of cell behaviour during morphogenesis. □

Methods

Time-lapse confocal analysis of cell behaviour

We injected embryos dorsally with 1 ng of *Xdd1*, *Xdsh* or control β -galactosidase messenger RNA and 100 pg of memEGFP mRNA (the ras membrane-localization (CAAX) sequence fused to the carboxy terminus of GFP)¹⁷. Open-face Keller explants were made at stage 10.5 (Fig. 1a); care was taken to remove head mesoderm and involuted axial mesoderm. We observed deep DMZ cells in the vegetal alignment zone⁸ by time-lapse confocal microscopy. Protrusive activity was quantified by counting new protrusions extended, existing protrusions withdrawn, or stable protrusions (present in both the first frame and the last frame of the movie). Orientation of lamellipodia was based on 90° sectors from the middle of the cell body (Fig. 1b). LWRs and angle of long axes were measured using NIH image.

Convergent extension in Keller explants

Embryos were injected dorsally with 1–1.5 ng of the mRNAs indicated and 75 pg of GFP. Explants were cut to an initial length of 1 mm. At stage 19, we measured and photographed GFP-positive explants. In Fig. 3a, the mean difference in length of uninjected control explants between stages 10.5 and 19 defines 100% elongation. Values for elongation of experimental explants are expressed as percentage of the control value; data shown are means of two or more experiments.

Xdsh localization

Embryos were injected with 60 pg of *Xdsh*–GFP¹⁴ mRNA and explants were prepared as above but cultured with 0.01 mg ml^{−1} FM 4-64 to label cell membranes. Animal caps were removed at stage 9 and cultured in 1 × MMR¹² ± activin. Samples were scanned simultaneously with FITC (Fig. 5a and b, left, and c, d) and Texas Red (Fig. 5a and b, right) filter sets. The low dose of *Xdsh*–GFP injected had no effect on development.

Received 13 January; accepted 24 March 2000.

- Sokol, S. Y. Analysis of Dishevelled signalling pathways during *Xenopus* development. *Curr. Biol.* **6**, 1456–1467 (1996).
- Axelrod, J. D., Miller, J. R., Shulman, J. M., Moon, R. T. & Perrimon, N. Differential recruitment of Dishevelled provides signaling specificity in the planar cell polarity and Wingless signaling pathways. *Genes Dev.* **12**, 2610–2622 (1998).
- Boutros, M., Paricio, N., Strutt, D. I. & Mlodzik, M. Dishevelled activates JNK and discriminates between JNK pathways in planar polarity and wingless signaling. *Cell* **94**, 109–118 (1998).
- Boutros, M. & Mlodzik, M. Dishevelled: at the crossroads of divergent intracellular signaling pathways. *Mech. Dev.* **83**, 27–37 (1999).
- Vogt, W. Gestaltungsanalyse am amphibienkeim mit örtlicher vitalfärbung. II. Teil. Gastrulation und mesodermbildung bei urodelen und anuran. *Wilhelm Roux Arch. EntwMech. Org.* **120**, 384–706 (1929).
- Keller, R., Shih, J., Wilson, P. & Sater, A. K. in *49th Symp. Soc. Develop. Biol. Cell–Cell Interactions in Early Development* (ed. Gerhart, J.) 93–107 (Wiley, New York, 1991).
- Keller, R., Shih, J. & Domingo, C. The patterning and functioning of protrusive activity during convergence and extension of the *Xenopus* organizer. *Development* (Suppl.) 81–91 (1992).
- Shih, J. & Keller, R. Patterns of cell motility in the organizer and dorsal mesoderm of *Xenopus laevis*. *Development* **116**, 915–930 (1992).
- Shih, J. & Keller, R. Cell motility driving mediolateral intercalation in explants of *Xenopus laevis*. *Development* **116**, 901–914 (1992).
- Cadigan, K. M. & Nusse, R. Wnt signaling: a common theme in animal development. *Genes Dev.* **11**, 3286–3305 (1997).
- Miller, J. R. *et al.* Establishment of the dorsal-ventral axis in *Xenopus* embryos coincides with the dorsal enrichment of dishevelled that is dependent on cortical rotation. *J. Cell Biol.* **146**, 427–437 (1999).
- Wallingford, J. B., Sater, A. K., Uzman, J. A. & Danilchik, M. V. Inhibition of morphogenetic movement during *Xenopus* gastrulation by injected sulfatase: implications for anteroposterior and dorsoventral axis formation. *Dev. Biol.* **187**, 224–235 (1997).
- Pierce, S. B. & Kimelman, D. Regulation of Spemann organizer formation by the intracellular kinase Xgsk-3. *Development* **121**, 755–765 (1995).
- Rothbächer, U. *et al.* Dishevelled phosphorylation, subcellular localization and homomerization regulate its role in early embryogenesis. *EMBO J.* **19**, 1010–1022 (2000).
- Howard, J. E. & Smith, J. C. Analysis of gastrulation: different types of gastrulation movement are induced by different mesoderm-inducing factors in *Xenopus laevis*. *Mech. Dev.* **43**, 37–48 (1993).
- Li, L. *et al.* Dishevelled proteins lead to two signaling pathways. Regulation of LEF-1 and c-Jun N-terminal kinase in mammalian cells. *J. Biol. Chem.* **274**, 129–134 (1999).

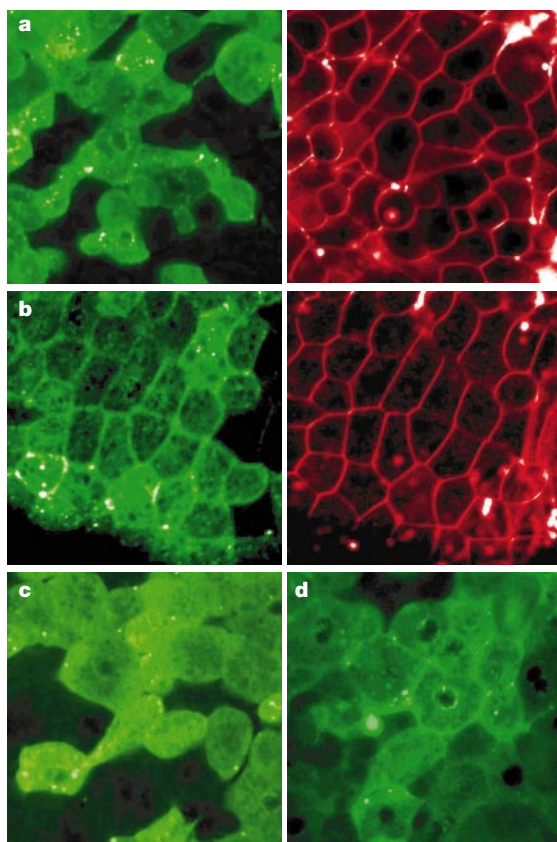


Figure 5 *Xdsh* is localized to the membrane in cells undergoing convergent extension. **a**, Left, *Xdsh*–GFP signal is found in a punctate, cytoplasmic distribution in posterior dorsal mesoderm cells that are not undergoing convergent extension at stage 10.75. **b**, Left, in anterior dorsal mesoderm cells engaged in convergent extension, *Xdsh*–GFP is localized to the cell membrane, consistent with signalling through the planar cell polarity cascade². **a**, **b**, Right, FM 4-64 membrane staining of the cells shown on the left. **c**, *Xdsh*–GFP is cytoplasmic in naive animal cap cells. **d**, *Xdsh*–GFP is present at the membrane in activin-treated animal cap cells undergoing convergent extension.

17. Moriyoshi, K., Richards, L. J., Akazawa, C., O'Leary, D. D. & Nakanishi, S. Labeling neural cells using adenoviral gene transfer of membrane-targeted GFP. *Neuron* **16**, 255–260 (1996).

Acknowledgements

We thank T. Grammer, M. Dionne, A. Sater, D. Keys, K. Liu and D. Frank for discussions, reading the manuscript, and blind scoring; S. Sokol for the Xdsh and Xdd1 plasmids; and D. Turner for the memEGFP plasmid. This work was supported by the NIH, NIMH and the Beckman Institute. J.B.W. was supported by NIH/NIGMS and ACS postdoctoral fellowships; B.A.R. by a US Department of Defense Breast Cancer Research Program Grant to C. Larabell; and K.M.V. by the Nathan and Violet David Scholars program at UC Berkeley.

Correspondence and requests for materials should be addressed to R.M.H. (e-mail: harland@socrates.berkeley.edu).

A receptor for phosphatidylserine-specific clearance of apoptotic cells

Valerie A. Fadok*, Donna L. Bratton*, David M. Rose*, Alan Pearson†, R. Alan B. Ezekewitz† & Peter M. Henson*

* Department of Pediatrics, National Jewish Medical and Research Center, 1400 Jackson Street, Denver, Colorado 80206, USA

† Massachusetts General Hospital, Boston, Massachusetts 02115, USA

The culmination of apoptosis *in vivo* is phagocytosis of cellular corpses. During apoptosis, the asymmetry of plasma membrane phospholipids is lost, which exposes phosphatidylserine externally^{1–4}. The phagocytosis of apoptotic cells can be inhibited stereospecifically by phosphatidylserine and its structural analogues, but not by other anionic phospholipids, suggesting that phosphatidylserine is specifically recognized^{1,5–10}. Using phage display, we have cloned a gene that appears to recognize phosphatidylserine on apoptotic cells. Here we show that this gene, when transfected into B and T lymphocytes, enables them to recognize and engulf apoptotic cells in a phosphatidylserine-specific manner. Flow cytometric analysis using a monoclonal antibody suggested that the protein is expressed on the surface of macrophages, fibroblasts and epithelial cells; this antibody, like phosphatidylserine liposomes, inhibited the phagocytosis of apoptotic cells and, in macrophages, induced an anti-inflammatory state. This candidate phosphatidylserine receptor is highly homologous to genes of unknown function in *Caenorhabditis elegans* and *Drosophila melanogaster*, suggesting that phosphatidylserine recognition on apoptotic cells during their removal by phagocytes is highly conserved throughout phylogeny.

Several potential candidates for phosphatidylserine (PS) recognition on apoptotic cells have been put forth, including CD36, CD68, CD14 and LOX-1 (refs 11–14). In addition, β 2GP1 may enhance uptake by bridging PS on the apoptotic cell to receptors on macrophages¹⁵. Several observations suggest, however, that receptors other than these scavenger/pattern-recognition molecules must exist. These molecules do not appear to discriminate between PS and other anionic phospholipids including phosphatidylinositol (PI)^{14,16–19}, whereas uptake of apoptotic cells by PS-recognizing macrophages is not blocked by phosphatidylinositol or other anionic phospholipids^{1,7}.

We generated monoclonal antibodies against human macrophages (HMDM) treated with transforming growth factor- β (TGF- β) and β -glucan to induce PS recognition⁷. Two IgM antibodies bound more strongly to stimulated HMDM than to unstimulated HMDM (Fig. 1a), but binding was inhibited by pre-incubating the macrophages with PS liposomes at 4°C before

staining (Fig. 1b; $n = 4$; $P = 0.0008$). Liposomes containing PI or pure phosphatidylcholine (PC) (data not shown) had no effect. Binding was also reduced by downregulating the antigen with PS liposomes for 30 min at 37°C (data not shown). Both antibodies inhibited the uptake of apoptotic cells by stimulated HMDM, but not that by unstimulated HMDM (Fig. 1c).

We then used monoclonal antibody 217 to examine adherent and suspension cells by flow cytometry. The binding of macrophages to mAb 217 was not species specific: mouse bone-marrow-derived macrophages only bound if they were stimulated to recognize PS⁶, whereas thioglycollate-elicited macrophages that constitutively

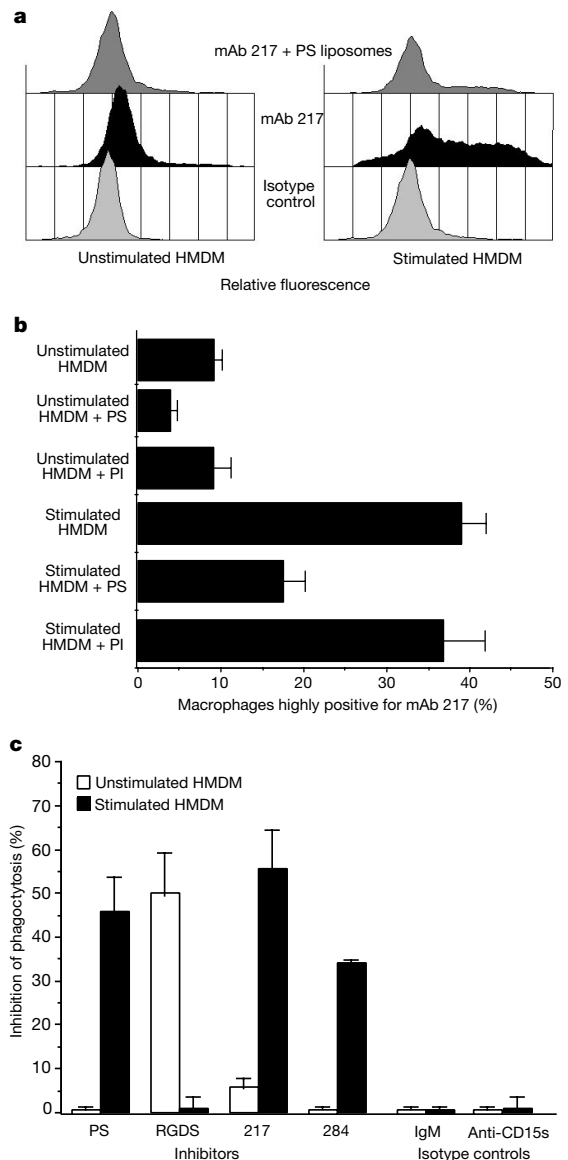


Figure 1 Monoclonal antibody 217 identifies a surface receptor expressed at high levels on human macrophages that recognize phosphatidylserine on apoptotic cells. **a**, Binding of mAb 217 to unstimulated and stimulated (TGF- β / β -glucan) human macrophages. Results represent one of four separate experiments, using macrophages from four different donors. **b**, Liposomes containing PS but not PI inhibit the binding of mAb 217 to stimulated macrophages ($n = 4$; $P = 0.0008$). **c**, Monoclonal antibodies 217 and 284 inhibit uptake of apoptotic cells by stimulated but not unstimulated HMDM. For these experiments, $48.9 \pm 8.6\%$ unstimulated macrophages contained apoptotic bodies and $55.2 \pm 6.1\%$ stimulated macrophages contained apoptotic bodies ($n = 5$; bars show mean $\pm$ s.e.m.). Liposomes containing PI and the tetrapeptide control RGES had no effect on uptake (not shown).

recognize PS^{1,5,8} bound. Also positive for expression were the mouse macrophage cell lines J774 and RAW264.7, as well as macrophages derived *in vitro* from embryonic stem cells. Monoclonal antibody 217 failed to bind circulating human blood cells (monocytes, lymphocytes, neutrophils, red blood cells) or cell lines mimicking them, including the lymphocytic lines, Jurkat T cells and mouse M12.C3 cells, and the myelocytic lines HL-60, PLB985, THP-1 and U937; treatment of the myelocytic, but not lymphocytic, cell lines with phorbol myristate acetate (PMA) induced expression of the antigen (data not shown). Several fibroblastic and epithelial cell lines (including NIH3T3, Swiss3T3, COS-1, COS-7, CHO-K1, HeLa, HC-11, HT1080, HEK293), as well as primary mouse lung fibroblasts and primary mouse mammary epithelial cells, were positive for binding to mAb 217.

Inhibition of apoptotic cell uptake by mAb 217 corresponded well with inhibition by PS liposomes. Thus, uptake by β -glucan-stimulated bone marrow macrophages, thioglycollate-elicited peritoneal macrophages, PMA-stimulated THP-1 cells, primary lung fibroblasts, 3T3 fibroblasts (data not shown) and the mammary epithelial cell line HC-11 was inhibited by PS liposomes and by mAb 217 to similar degrees (Fig. 2a).

We carried out western blots using mAb 217 on whole-cell extracts from human Jurkat T lymphocytes, mouse M12.C3 B lymphocytes, human THP-1 cells differentiated into macrophages using phorbol ester, human HT1080 fibrosarcoma cells and mouse 3T3 fibroblasts. A band of relative molecular mass $\sim 70,000$ ($M_r \approx 70K$) was observed in only those cell lines that showed surface expression of the antigen for mAb 217 (Fig. 2b).

We cloned the antigen for mAb 217 by using phage display and biopanning with mAb 217. The sequence obtained from the clones was used to search the Genbank databases. One significant match was made to a complementary DNA from a human brain library for a protein (KIAA 0585) of unknown function highly homologous to an uncharacterized open reading frame (ORF) in *C. elegans*²⁰. Polymerase chain reaction with reverse transcription (RT-PCR) previously determined that the expression of this transcript was high in the heart, skeletal muscle and kidney, and moderate or low in brain, placenta, lung, liver, pancreas, spleen, thymus, prostate, testis and ovary²⁰. A number of homologous sequences were present in human, mouse, *Drosophila* and *C. elegans* expressed sequence tag (EST) databases. An EST clone was obtained (accession number R24727) through the IMAGE (Integrated Molecular Analysis of

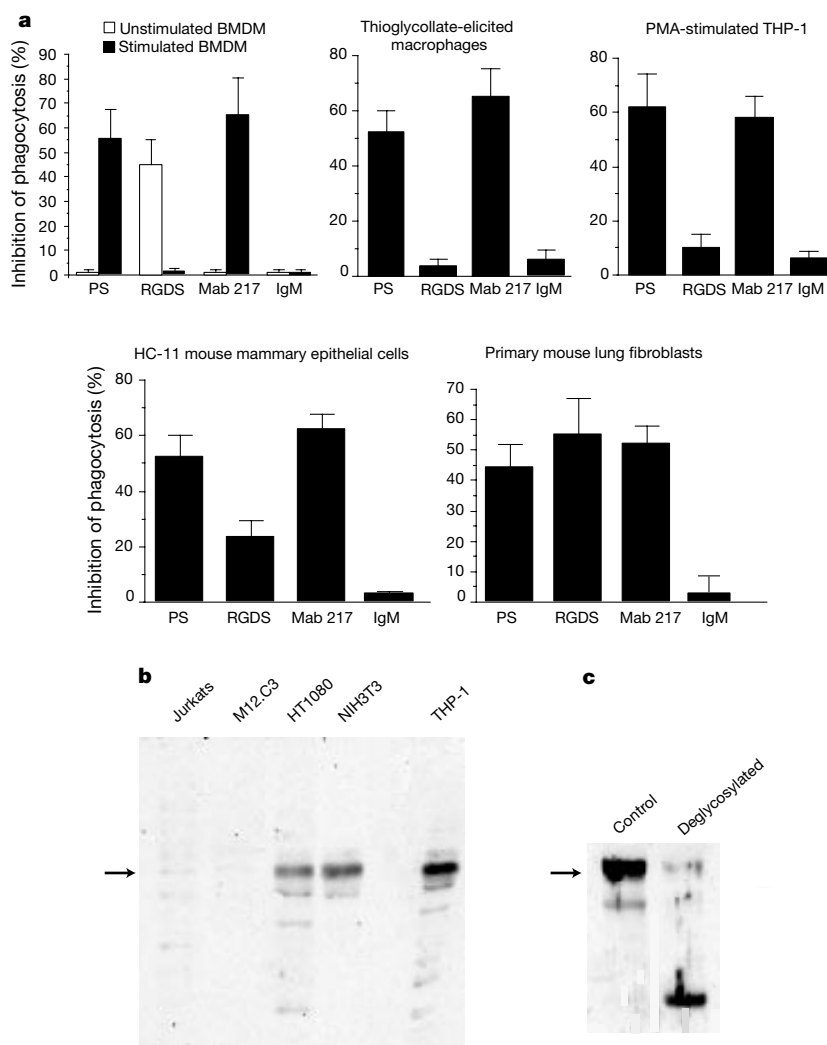
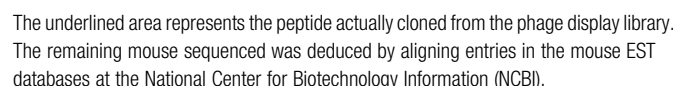


Figure 2 Monoclonal 217 inhibits uptake of apoptotic Jurkat T cells by several phagocytes. Mouse bone-marrow-derived macrophages were stimulated to induce the ability to recognize PS⁶. The mean percentages of phagocytes containing apoptotic bodies were as follows: unstimulated BMDM, 45%; stimulated BMDM, 42%; thioglycollate-elicited macrophages, 62%; THP-1 cells, 24%; HC-11 cells, 15%; primary lung

fibroblasts, 17.8% ($n = 5$; bars show mean $\pm$ s.e.m.) **b**, Western blot using mAb 217 reveals a 70K band only in those cells which show surface expression of the antigen by flow cytometry. **c**, Deglycosylation of the candidate protein with *O*-glycosidases shifts the molecular weight from 70K to $\sim 50K$.

The predicted molecular weight for this protein, based on sequence analysis, was 47–48K for the human, mouse and *Drosophila* gene products, and 40K for that of the nematode. Because western blotting suggested an apparent molecular weight of 70K, it was important to determine whether deglycosylation would reduce the apparent size. The predicted extracellular domain contains a run



of serines compatible with the possibility that this protein could be O-glycosylated (Fig. 3). Extracts from HT1080 and NIH3T3 cells were deglycosylated using enzymes, the proteins separated by polyacrylamide gel electrophoresis (PAGE), and a western blot carried out using mAb 217. Treatment with glycosidases resulted in a shift from 70K to ~50K, compatible with the predicted molecular weight of 47K (Fig. 2c).

Because the adherent cell lines one would normally use for transfection (COS-7, COS-1, CHO-K1, HEK 293) already showed surface expression of the antigen for mAb 217, we selected two unambiguously negative lymphocyte lines: the mouse Class II negative B cell line M12.C3 and the human Jurkat T cell. Neither is capable of binding or phagocytosing apoptotic cells. Using flow cytometry, we showed that transiently transfected mouse M12.C3

cells (Fig. 4a) and human Jurkat T cells (data not shown) bound to mAb 217. Transiently transfected M12.C3 cells were used to examine the function of this receptor in mediating the binding and phagocytosis of apoptotic cells. About 25% of M12.C3 cells were able to bind mAb 217 and apoptotic targets, and ~12.5% were able to engulf apoptotic cells (Fig. 4b, upper panel). When the transfected cells were sorted into those positive for mAb 217 binding and those negative, we found that only the positive cells were able to bind and engulf apoptotic cells (Fig. 4b, lower panel). M12.C3 cells transfected with the empty vector showed little binding to and absolutely no phagocytosis of apoptotic cells, despite expressing low levels of $\alpha\text{v}\beta 3$, an integrin incriminated in uptake of apoptotic cells²³. We needed to confirm that the putative PS receptor could mediate uptake of apoptotic cells by Jurkat T cells, as they do not

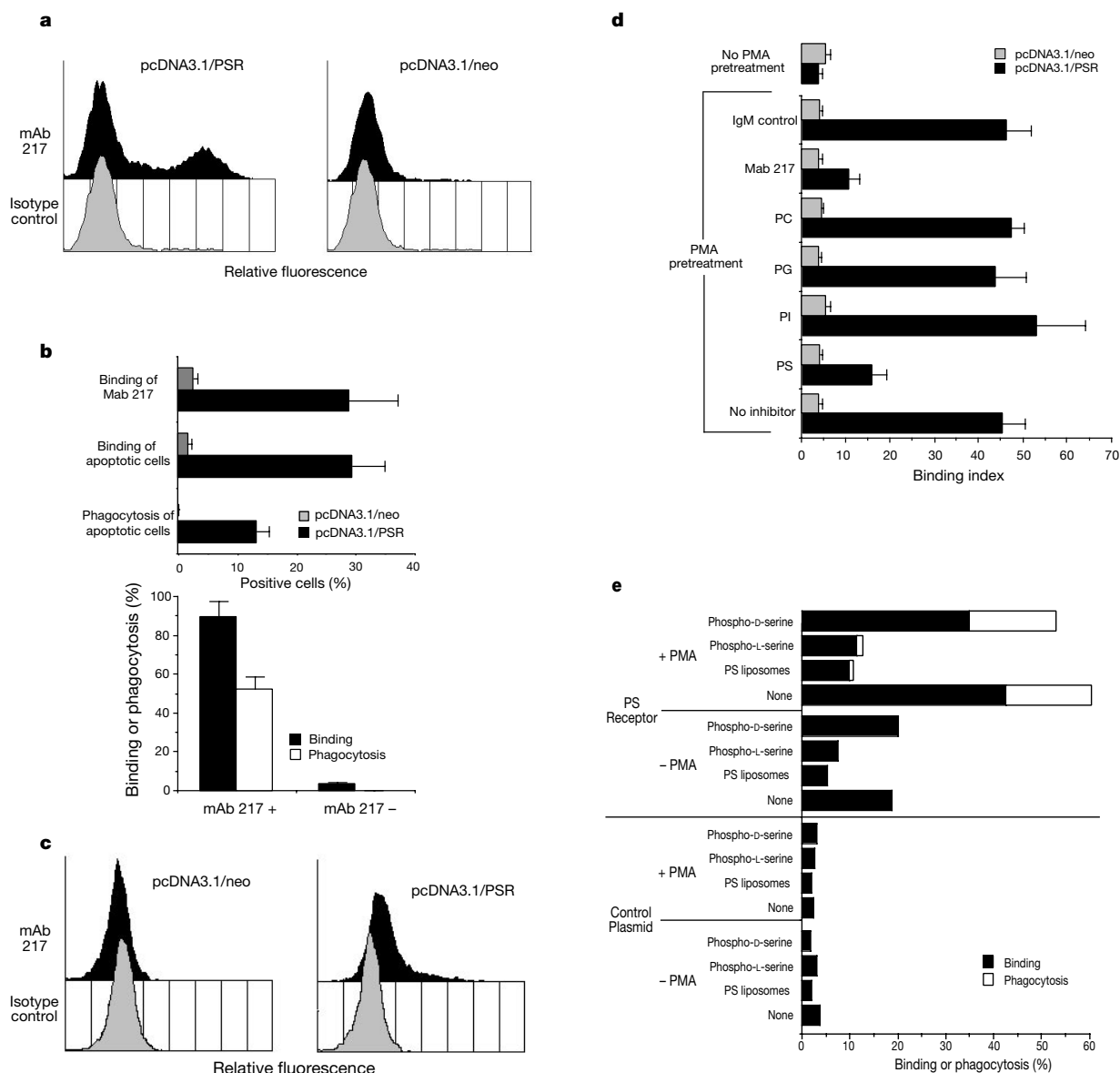


Figure 4 Transfection of lymphocytes with the candidate PS receptor induces PS-specific binding and uptake of apoptotic cells. **a**, Monoclonal antibody 217 binds to M12.C3 cells transiently transfected with the candidate PS receptor (pcDNA3.1/PSR), but not with control plasmid (pcDNA3.1/neo). Results represent one of four experiments. **b**, M12.C3 B cells transiently transfected with the candidate PSR phagocytose bind to and phagocytose apoptotic cells ($n = 4$; bars show mean $\pm$ s.e.m.). Upper panel, unsorted cells. Lower panel, cells sorted by flow cytometry into mAb 217 positive or negative. **c**, Binding of mAb 217 to Jurkat T cells stably transfected with the candidate PS receptor. Results represent

10 individual experiments, and show the average expression seen after stimulation. Without PMA stimulation, surface expression of the receptor is very low. **d**, Apoptotic cells bind to stably transfected Jurkat T cells in a PS-dependent manner, which is inhibited by mAb 217 ($n = 5$; bars show mean $\pm$ s.e.m.). Binding index represents the average per cent of transfectants binding apoptotic cells $\times$ average number apoptotic cells per transfectants (1.1). **e**, Binding and phagocytosis of symmetric red cell ghosts by Jurkat T cells stably transfected with candidate PSR are inhibited stereospecifically by phospho-L-serine and not phospho-D-serine ($n = 3$).

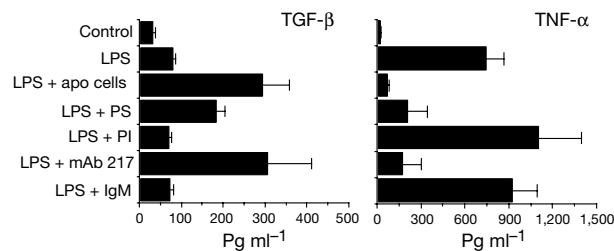


Figure 5 Treatment of J774 mouse macrophages with mAb 217, PS-containing liposomes or apoptotic cells stimulates TGF- β secretion and inhibits LPS-induced TNF- α production. Pg, picograms ml⁻¹.

express any of the known surface receptors for uptake of apoptotic cells, including $\alpha v \beta 3$. After successful transient transfections, Jurkat T cells could bind to and engulf apoptotic cells. To overcome the problems of transient transfection, the Jurkat T cells were stably transfected by selection in G418. Four lines were obtained that consistently expressed low levels of the receptor after stimulation with 160 nM PMA (Fig. 4c).

The stable Jurkat T cells bound reliably to apoptotic cells only when PS receptor expression was increased with PMA (Fig. 4d). Monoclonal antibody 217 and PS liposomes inhibited binding; control IgM and other anionic phospholipids did not (Fig. 4d). These results exactly recapitulated our observations with macrophages that recognize PS^{1,7}. In addition, $5.8 \pm 1.2\%$ of the stably transfected cells phagocytosed apoptotic cells; phagocytosis was never observed in Jurkat cells transfected with the control plasmid (for example, Fig. 4c). We also found that Jurkat T cells transfected with the candidate PSR could bind to symmetric red cell ghosts, and that binding was inhibited by PS liposomes and by phospho-L-serine, not phospho-D-serine, showing the same stereospecificity seen with macrophages (Fig. 4e). If, however, the transfectants were pretreated with PMA to increase surface expression of the receptor, their ability to bind symmetric red cells was increased and they engulfed them as well.

Uptake of apoptotic cells by macrophages is associated with the production of TGF- β and the downregulation of inflammatory cytokines^{24,25}. Because phosphatidylserine downregulates several macrophage functions, including tumour necrosis factor- α (TNF- α) production²⁶, we wanted to determine whether the candidate PS receptor could provide the signal for this cytokine response. PS-containing liposomes and mAb 217 stimulated TGF- β and decreased TNF- α production in LPS-stimulated J774 macrophages (Fig. 5), indicating that the candidate PS receptor may have an important function in the downregulation of inflammatory responses after uptake of apoptotic cells by macrophages. This antibody was also able to stimulate TGF- β production from NIH3T3 cells and transfected Jurkat T lymphocytes (data not shown); their engulfment of apoptotic cells was also inhibited by PS and by mAb 217.

The high homology between the mammalian protein and genes of unknown function in *C. elegans* and *D. melanogaster* implies that PS recognition may have been preserved across phylogeny. Three of the six genes known to be involved in phagocytosis of apoptotic bodies in *C. elegans* are homologous to mammalian proteins^{27–29}, suggesting that the basic machinery mediating phagocytosis may be conserved in most multicellular organisms. Clearance of apoptotic bodies is also a feature of apoptotic cell death in *Drosophila*³⁰; for example, croquemort, a homologue of mammalian CD36, has a critical function³⁰. It will be important to determine whether our candidate PS receptor also mediates uptake in *Drosophila* and *C. elegans*.

On the basis of the data shown here, we propose that this gene product is a good candidate for a PS-specific receptor, although we cannot rule out at this time that it facilitates PS recognition by some other function which does not involve direct binding to PS.

Nevertheless, we believe this protein is critical in mediating uptake of apoptotic cells and the anti-inflammatory effects of clearance in mammals. □

Methods

Cells and cell lines

Human monocytes were obtained from normal donors and cultured in X-Vivo 10 (BioWhittaker) containing 10% human serum. The differentiated macrophages (HMDM) were used at 7 d of culture. Mouse bone-marrow-derived macrophages were obtained by culturing bone marrow cells in M-CSF for 5–7 d. All other cell lines, with the exception of M12.C3 mouse lymphocytes and the HC-11 mouse mammary epithelial cells, were obtained from the American Type Culture Collection (Manassas, VA).

Development of monoclonal antibodies

HMDM stimulated to induce PS receptor expression⁷ were injected four times at monthly intervals intraperitoneally into BALB/c mice. The sera was tested for ability to bind at higher titre to stimulated versus unstimulated HMDM using whole-cell ELISA. Supernatants resulting from the fusion were tested by whole-cell ELISA for stimulated macrophages, by flow cytometry for unstimulated macrophages and stimulated macrophages pre-incubated (or not) with 100 μ M of PS or PI liposomes (containing 50:50 molar ratios of PS or PI to PC)¹ for 30 min at 4 °C. Two hybridoma supernatants (mAb 217, mAb 284, IgM, kappa chain) bound at higher levels to stimulated macrophages and were inhibited from binding by PS liposomes, but not PI or PC liposomes.

Flow cytometry, cell sorting

Roughly $0.5-1 \times 10^6$ cells were incubated in 100 μ l of HBSS containing 2% FCS and 100 μ l of hybridoma supernatant. Isotype controls included anti-CD15s (PharMingen), a mouse IgM which bound at high levels to human macrophages, and mouse myeloma IgM (Calbiochem). The secondary antibody was Cy-3 conjugated goat anti-mouse IgM, μ -chain-specific (F(ab')₂, Jackson Immunoresearch Laboratories). Stained cells were analysed using a Becton-Dickson FacsCAN cytometer. Data were analysed using PCLysis software. For the transfected cells, purified mAb 217 was used as the primary reagent at 100–200 μ g ml⁻¹. Adherent cells were prepared for flow cytometry by collecting them at 4 °C in Hank's BSS containing 5 mM EDTA for 15 min, followed by gentle scraping. Trypsin was not used, as it removes the antigen from the cell surface. Transiently transfected M12.C3 were sorted into mAb 217 positive or negative using the Mo-Flow cell sorter.

Western blotting

Cells were lysed in 50 mM Tris pH 7.4 containing 150 mM NaCl, 1% Triton-X 100, 0.25% deoxycholate and a protease inhibitor cocktail (AEBSE, pepstatin A, E-64, bestatin, leupeptin, aprotinin; Sigma); proteins were separated by SDS-PAGE and transferred to nitrocellulose. After a 1-h wash in 4 M urea, 50 mM NaCl and 5 mM EDTA, membranes were blocked with 5% dried milk, incubated with mAb 217 at 3 μ g ml⁻¹ overnight at 4 °C, treated with peroxidase-labelled goat anti-mouse μ -chain-specific F(ab')₂ for 1 h (Jackson Immunoresearch), and developed by chemiluminescence (ECL, Amersham Pharmacia Biotech). Whole-cell extracts were deglycosylated with NANase II, O-glycosidase DS, HEXase I and GALase III following denaturation according to protocols recommended by Bio-Rad.

Cloning of the receptor

A directional cDNA library was made using the T7 Select system available from Novagen. First, mRNA was isolated from J774 cells using Fast-Track (Invitrogen). cDNA was generated using directional HindIII random primers (Novagen) by reverse transcription using MMLV reverse transcriptase. The ends were flushed with T4 DNA polymerase and directional EcoRI/Hind III linkers added. The sample was digested with EcoRI and HindIII and ligated into the T7 Select 1-1b vector (Novagen), and amplified using the plate lysate amplification method recommended for this system. Dishes (100 mm³) were coated with purified mAb 217, an IgM control (anti-TNP, Pharmingen) or BSA, and blocked with 1% BSA and 5% dry milk. Viral lysates (2 ml) were added to the plates, which were incubated at room temp for 2 h, and then thoroughly washed with TBST. The bound virus particles were removed using 1% SDS, and 250 μ l was added to 50 ml of the bacterial host strain

(BLT5615) for lysis again. Lysis was seen only in material biopanned from plates coated with mAb 217; materials from plates coated with nonspecific IgM or BSA were not lytic. After 5 rounds of biopanning, plaques were isolated and 12 were selected for sequencing using the ABI PRISM dRhodamine cycle sequencing system (PE Biosystems). Sequences were analysed and aligned using MacVector 6.0 (Oxford Molecular Group); six were identical. Digoxigenin-labelled probes (DIG System, Boehringer Mannheim) were made from the six identical sequences and used to probe mRNA derived from J774 cells by northern analysis. A 2-kb band was observed, confirming that the clones were derived from J774.

The full mouse sequence was obtained by aligning multiple entries in the mouse EST databases (Genbank). The sequence for the human homologue KIAA 0585 (ref. 20) was available in Genbank (accession number ABO11157). The *C. elegans* homologue can be found in cosmid F29B9 (U70849.1) as accession number AAB09116. The complete sequence of the *D. melanogaster* homologue was determined by sequencing a cDNA clone L0022859, whose partial sequence was already known (Genbank accession number AA940716). The clone was obtained from the Berkeley Drosophila Genome Project (BDGP/HHMI EST project) and Research Genetics.

Transfection of M12.C3 and Jurkat T cells

A human EST clone (accession number R24727) containing the full-length gene was obtained through the IMAGE consortium via ATCC (Manassas, VA). The insert subcloned into pcDNA 3.1/neo at *HindIII*/*NotI* sites (Invitrogen). M12.C3 cells were transfected in OptiMEM 1 medium using Lipofectamine Plus (both from GibcoBRL Life Science Technologies). Jurkat T cells were transfected in OptiMEM 1 medium using DMRIE C. Jurkat T cells were stimulated with PHA alone ($1 \mu\text{g ml}^{-1}$) or with PMA (1.6 nM) and PHA to induce expression of the receptor. After 48 h, the cells were analysed for expression of the PS receptor using mAb 217 and flow cytometry. For stable transfections, Jurkat T cells were selected in G418 (1 mg ml^{-1}). Surface expression of the receptor was induced by treatment of the stably transfected cells with 160 nM PMA in the absence of G418 for 48 h. The candidate PSR was also inserted into pRC/RSV (PharMingen), or pUP (provided by B. Schaefer), and transfected into Jurkat T cells or M12.C3 cells by electroporation.

Binding and phagocytosis of apoptotic cells

Macrophages, fibroblasts or epithelial cells were pre-incubated with 100 μM liposomes (containing 50:50 molar ratio of PS:PC or PI:PC), 1 mM RGDS or RGES, undiluted hybridoma supernatants containing mAb 217 or mAb 284 (for HMDM), or 100 $\mu\text{g ml}^{-1}$ purified mAb 217 or 100 $\mu\text{g ml}^{-1}$ isotype controls for 30 min before adding apoptotic Jurkat T cells at ratio of 5:1 target:phagocyte. Apoptosis was induced in Jurkat T cells by ultraviolet irradiation as described²⁴. THP-1 cells were stimulated with PMA (160 nM) for 72 h to induce a macrophage phenotype, the ability to adhere and the ability to take up apoptotic cells. After 1 h, the phagocytes were washed and fixed for light microscopic evaluation of uptake. Apoptotic Jurkat T cells, HL-60 cells or human neutrophils (induced by ultraviolet irradiation)^{24,25} were added to the transfected M12.C3 or Jurkat T cells at ratio of 5:1 target to transfectant and incubated for 1 h to assess binding in the presence or absence of mAb 217 (100–200 $\mu\text{g ml}^{-1}$), IgM controls (100–200 $\mu\text{g ml}^{-1}$), or 100 μM liposomes containing PC, or 50:50 molar ratios of PS:PC, PI:PC, or PG:PC (all phospholipids obtained from Avanti Polar Lipids). Aliquots were removed and assessed for binding and phagocytosis by light microscopy as described^{15–17}. For phagocytosis, the stably transfected Jurkat T cells were immobilized on glass slides treated with poly-L-lysine (100 $\mu\text{g ml}^{-1}$). Apoptotic cells were added for 2 h, and then the slides were washed, fixed with methanol and stained with a modified Wright's Giemsa stain.

Binding and phagocytosis of symmetric red cell ghosts

Human red blood cells were lysed in 1:50 PBS containing 1 mM CaCl_2 on ice for 1 h, and rescaled at 37 °C for 30 min by addition of 5×PBS. Exposure of phosphatidylserine was confirmed by binding of annexin-FITC, as assessed by flow cytometry.

Cytokine analysis

J774 cells were treated with 10 ng ml^{-1} LPS in X-Vivo 10 medium and either apoptotic Jurkat T cells, 100 μM liposomes containing 50 molar % PS or PI, 100 $\mu\text{g ml}^{-1}$ mAb 217, or 100 $\mu\text{g ml}^{-1}$ mouse IgM isotype control added. Supernatants were evaluated for cytokine production by ELISA 18 h later²⁴. 3T3 fibroblasts and Jurkat T cells stably transfected with the candidate gene were treated with mAb 217 or isotype control and assessed for TGF- β production 18 h later.

Statistical analysis

Statistical analysis (analysis of variance (ANOVA) and Tukey–Kramer HSD) was carried out using JMP software (SAS Institute).

Received 3 February; accepted 7 March 2000.

1. Fadok, V. A. *et al.* Exposure of phosphatidylserine on the surface of apoptotic lymphocytes triggers specific recognition and removal by macrophages. *J. Immunol.* **148**, 2207–2216 (1992).
2. Martin, S. J. *et al.* Early redistribution of plasma membrane phosphatidylserine is a general feature of apoptosis regardless of the initiating stimulus: inhibition by overexpression of Bcl-2 and Abl. *J. Exp. Med.* **182**, 1545–1556 (1995).
3. Verhoven, B., Schlegel, R. A. & Williamson, P. Mechanisms of phosphatidylserine exposure, a phagocyte recognition signal, on apoptotic T lymphocytes. *J. Exp. Med.* **182**, 1597–1601 (1995).
4. van den Eijnde, S. M. *et al.* Phosphatidylserine exposure by apoptotic cells is phylogenetically conserved. *Apoptosis* **3**, 9–16 (1998).

5. Fadok, V. A. *et al.* Different populations of macrophages use either the vitronectin receptor or the phosphatidylserine receptor to recognize and remove apoptotic cells. *J. Immunol.* **149**, 4029–4035 (1992).
6. Fadok, V. A. *et al.* Particle digestibility is required for induction of the phosphatidylserine recognition mechanism used by murine macrophages to phagocytose apoptotic cells. *J. Immunol.* **151**, 4274–4285 (1993).
7. Fadok, V. A., Warner, M. L., Bratton, D. L. & Henson, P. M. CD36 is required for phagocytosis of apoptotic cells by human macrophages that use either a phosphatidylserine receptor or the vitronectin receptor ($\alpha\text{v}\beta 3$). *J. Immunol.* **161**, 6250–6257 (1998).
8. Pradhan, D., Krahling, S., Williamson, P. & Schlegel, R. A. Multiple systems for recognition of apoptotic lymphocytes by macrophages. *Mol. Biol. Cell.* **8**, 767–778 (1997).
9. Bennett, M. R., Gibson, D. F., Schwartz, S. M. & Tait, J. F. Binding and phagocytosis of apoptotic vascular smooth muscle cells is mediated in part by exposure of phosphatidylserine. *Circ. Res.* **77**, 1136–1142 (1995).
10. Shiratsuchi, A., Umeda, M., Ohba, Y. & Nakanishi, Y. Recognition of phosphatidylserine on the surface of apoptotic spermatogenic cells and subsequent phagocytosis by Sertoli cells of the rat. *J. Biol. Chem.* **272**, 2354–2358 (1997).
11. Savill, J., Hogg, N., Ren, Y. & Haslett, C. Thrombospondin cooperates with CD36 and the vitronectin receptor in macrophage recognition of neutrophils undergoing apoptosis. *J. Clin. Invest.* **90**, 1513–1522 (1992).
12. Sambrano, G. R. & Steinberg, D. Recognition of oxidatively damaged and apoptotic cells by an oxidized low density lipoprotein receptor on mouse peritoneal macrophages: role of membrane phosphatidylserine. *Proc. Natl Acad. Sci. USA* **92**, 1396–1400 (1995).
13. Devitt, A. *et al.* Human CD14 mediates recognition and phagocytosis of apoptotic cells. *Nature* **392**, 505–509 (1998).
14. Oka, K. *et al.* Lectin-like oxidized low-density lipoprotein receptor 1 mediates phagocytosis of aged/apoptotic cells in endothelial cells. *Proc. Natl Acad. Sci. USA* **95**, 9535–9540 (1998).
15. Balasubramanian, K., Chandra, J. & Schroit, A. J. Immune clearance of phosphatidylserine-expressing cells by phagocytes. The role of $\beta 2$ -glycoprotein I in macrophage recognition. *J. Biol. Chem.* **272**, 31113–31117 (1997).
16. Rigotti, A., Acton, S. L. & Krieger, M. The class B scavenger receptors SR-BI and CD36 are receptors for anionic phospholipids. *J. Biol. Chem.* **270**, 16221–16224 (1995).
17. Wang, P. Y., Kitchens, R. L. & Munford, R. S. Phosphatidylserine bind to plasma membrane CD14 and can prevent monocyte activation by bacterial lipopolysaccharide. *J. Biol. Chem.* **273**, 24309–24313 (1998).
18. Ryeom, S. W., Sparrow, J. R. & Silverstein, R. L. CD36 participates in the phagocytosis of rod outer segments by retinal pigment epithelium. *J. Cell. Sci.* **109**, 387–395 (1996).
19. Shiratsuchi, A., Kawasaki, Y., Ikemoto, M., Arai, H. & Nakanishi, Y. Role of class B scavenger receptor type 1 in phagocytosis of apoptotic rat spermatogenic cells by Sertoli cells. *J. Biol. Chem.* **274**, 5901–5908 (1999).
20. Nagase, T. *et al.* Prediction of the coding sequences of unidentified human genes. IX. The complete sequences of 100 new cDNA clones from brain which can code for large proteins *in vitro*. *DNA Res.* **5**, 31–39 (1998).
21. Heijne, G. Membrane protein structure prediction, hydrophobicity analysis, and the positive-inside rule. *J. Mol. Biol.* **225**, 487–494 (1992).
22. Igarashi, K. *et al.* A novel phosphatidylserine binding peptide motif defined by an anti-idiotypic monoclonal antibody. Localization of phosphatidylserine-specific binding sites on protein kinase C and phosphatidylserine decarboxylase. *J. Biol. Chem.* **270**, 29075–29078 (1995).
23. Savill, J., Dransfield, I., Hogg, N. & Haslett, C. Vitronectin receptor-mediated phagocytosis of cells undergoing apoptosis. *Nature* **343**, 170–173 (1990).
24. Fadok, V. A. *et al.* Macrophages that have ingested apoptotic cells *in vitro* inhibit proinflammatory cytokine production through autocrine/paracrine mechanisms involving TGF- β , PGE $_2$, and PAF. *J. Clin. Invest.* **101**, 890–898 (1998).
25. McDonald, P. P., Fadok, V. A., Bratton, D. L. & Henson, P. M. Transcriptional and translational regulation of inflammatory mediator production by endogenous TGF- β in macrophages that have ingested apoptotic cells. *J. Immunol.* **163**, 6164–6172 (2000).
26. Aramaki, Y., Matsuno, R., Nitta, F., Arima, H. & Tsuchiya, S. Negatively charged liposomes inhibit tyrosine phosphorylation of 41-kDa protein in murine macrophages stimulated with LPS. *Biochem. Biophys. Res. Commun.* **231**, 827–830 (1997).
27. Wu, Y. C. & Horvitz, H. R. *C. elegans* phagocytosis and cell-migration protein CED-5 is similar to human DOCK180. *Nature* **392**, 501–504 (1998).
28. Wu, Y. C. & Horvitz, H. R. The *C. elegans* cell corpse engulfment gene *ced-7* encodes a protein similar to ABC transporters. *Cell* **93**, 951–960 (1998).
29. Liu, Q. A. & Hengartner, M. O. Candidate adaptor protein CED-6 promotes the engulfment of apoptotic cells in *C. elegans*. *Cell* **93**, 961–972 (1998).
30. Franc, N. C., Heitzler, P., Ezekowitz, R. A. & White, K. Requirement for croquemort in phagocytosis of apoptotic cells in *Drosophila*. *Science* **284**, 1991–1994 (1999).

Acknowledgements

We thank K. Christensen and D. Edwards for monoclonal antibody production; L. Remigio and D. Riches for providing mouse BMDM; M. Neville and B. Groener for the gift of HC-11 cells; J. Freed for valuable critical comments and a gift of M12.C3 B cells; C. Leslie for provision of PLB 985 cells; A. Stewart for help with culture of primary lung fibroblasts; B. Schaefer for technical advice and the gift of the pUP vector; J. Westcott for provision of the cytokine assays; B. Townend for assistance with flow cytometry and cell sorting; M. Jones for assistance with fluorescence microscopy; T. Konowal for generation of the cDNA library from J774 cells; P. Taylor for assistance with bioinformatics; R. Schlegel and P. Williamson for critical comments; and D. Voelker for technical advice and constructive critical commentary.

Correspondence and requests for materials should be addressed to VAF (e-mail: fadokv@njc.org). The human EST clone has been deposited in the GenBank under accession number R24727.

JNK is required for effector T-cell function but not for T-cell activation

Chen Dong*, Derek D. Yang*†, Cathy Tournier‡, Alan J. Whitmarsh‡, Jie Xu*, Roger J. Davis† & Richard A. Flavell*

* Howard Hughes Medical Institute, Section of Immunobiology,

Yale University School of Medicine, New Haven, Connecticut 06520, USA

‡ Howard Hughes Medical Institute, Program in Molecular Medicine, Department of Biochemistry and Molecular Biology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA

† Present address: Lilly Research Laboratories, Indianapolis, Indiana 46285, USA

The hallmark of T-cell activation is the production of interleukin 2 (IL-2). c-Jun amino-terminal kinase (JNK), a MAP kinase that phosphorylates c-Jun and other components of the AP-1 group of transcription factors^{1,2}, has been implicated in the activation of IL-2 expression^{3,4}. Previously, we found that T cells from mice deficient in the *Jnk1* or *Jnk2* gene can be activated and produce IL-2 normally, but are deficient in functional differentiation into Th1 or Th2 subsets^{5,6}. However, studies of mice with compound mutations indicate that JNK1 and JNK2 are redundant during mouse development⁷. Here we use three new mouse models in which peripheral T cells completely lack JNK proteins or signalling, to test whether the JNK signalling pathway is crucial for IL-2 expression and T-cell activation. Unexpectedly, these T cells made more IL-2 and proliferated better than wild-type cells. However, production of effector T-cell cytokines did require JNK. Thus, JNK is necessary for T-cell differentiation but not for naive T-cell activation.

As a first step towards resolving the role of JNK in T-cell activation, we expressed activated JNK1 (MKK7 plus JNK1) in Jurkat T cells. It is known that a minimal IL-2 promoter containing multimerized DNA elements that bind NFAT-AP1 (nuclear factor of activated T cells, activating protein 1) can be strongly activated by phorbol ester and calcium ionophore. We found, however, that expression of MKK7 plus JNK1 did not activate, but inhibited IL-2 NFAT-AP1 reporter activity (data not shown). Unexpectedly, these *in vitro* experiments indicated that the MKK7–JNK pathway does

not activate and may inhibit IL-2 expression. Also, we previously reported that the requisite components of the JNK signalling pathway (JNK1, JNK2, MKK4 and MKK7) are barely detectable in naive and recently activated T cells, but instead are highly induced by 48 h after activation, by which time IL-2 expression is downregulated⁸.

We were concerned that overexpression studies of this nature might not reflect the *in vivo* situation. Three genes encode the JNK protein kinases in mammals; *Jnk1* and *Jnk2* (but not *Jnk3*) are expressed in the immune system. To test whether the JNK signalling pathway contributes to IL-2 gene regulation *in vivo*, we bred transgenic mice that express dominant-negative JNK1 (*dnJNK1*) in T cells⁹ with *Jnk2*^{−/−} mice⁶. These *dnJNK1*⁺*Jnk2*^{−/−} mice developed B and T lymphocytes normally (data not shown), indicating that JNK1 and JNK2 are not essential for lymphocyte development. We isolated peripheral CD4⁺ T cells from these mice and treated them with ConA, or anti-CD3 with or without anti-CD28. IL-2 production was measured by enzyme-linked immunosorbent assay (ELISA). Surprisingly, CD4⁺ T cells from *dnJNK1*⁺*Jnk2*^{−/−} mice produced more IL-2 protein and IL-2 messenger RNA than control wild-type mice (Fig. 1a and b). These data also failed to support a role for JNK as a positive regulator of IL-2 gene expression and T-cell activation.

To further substantiate the observations made with *dnJNK1*⁺*Jnk2*^{−/−} CD4⁺ T cells, we generated primary embryonic stem (ES) cells from blastocysts of *Jnk1*^{+/−}*Jnk2*^{−/−} intermatings. ES cell clones with disruptions of both *Jnk1* and *Jnk2* were identified by Southern blot (Fig. 2a). The absence of JNK1 and JNK2 protein was confirmed by immunoblot (Fig. 2b) and measurement of c-Jun phosphorylation in response to ultraviolet radiation (Fig. 2c). No expression of the JNK3 isoform could be detected by immunoblot analysis of the *Jnk1*^{−/−}*Jnk2*^{−/−} ES cells (data not shown). Furthermore, reporter gene assays demonstrated that the *Jnk1*^{−/−}*Jnk2*^{−/−} ES cells were defective in the activation of the chimaeric transcription factor GAL4–cJun by MEKK1 and MLK3, known activators of the JNK signalling pathway (Fig. 2d). Together, these data demonstrate that the *Jnk1*^{−/−}*Jnk2*^{−/−} ES cells are structurally and functionally null for JNK. As JNK1 and JNK2, but not JNK3, are expressed by T cells, we used the *Jnk1*^{−/−}*Jnk2*^{−/−} ES cells to reconstitute recombination activating gene 1 (*Rag1*) knockout mice. Like the *dnJNK1*⁺*Jnk2*^{−/−} mice, *Rag1*^{−/−}*Jnk1*^{−/−}*Jnk2*^{−/−} mice had normal development of T and B lymphocytes, exhibiting normal thymocyte and peripheral lymphocyte

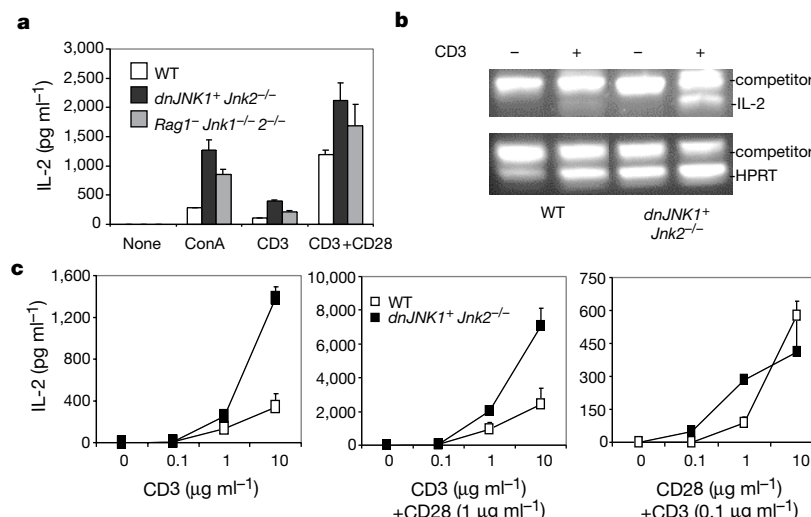


Figure 1 Increased IL-2 production by T cells with defects in the JNK signalling pathway. **a**, CD4⁺ T cells from 6–8-week-old wild-type (WT), *dnJNK1*⁺*Jnk2*^{−/−}, or *Rag1*^{−/−} mice reconstituted with *Jnk1*^{−/−}*Jnk2*^{−/−} ES cells were purified from spleen and lymph nodes and cultured with or without plate-bound anti-CD3 or anti-CD3 plus CD28 antibodies for

24 h. IL-2 production was assayed by ELISA. **b**, IL-2 mRNA was measured by competitive RT–PCR analysis. **c**, The CD4⁺ cells were treated with the indicated doses of anti-CD3 or anti-CD28 antibodies for 24 h and IL-2 production measured by ELISA.

populations (data not shown). Notably, the CD4⁺ T cells from these mice showed no defect in T-cell activation. They also had increased IL-2 production (compared with wild-type cells) following activation by ConA or anti-CD3 plus anti-CD28 (Fig. 1a), rather than the absence of IL-2 expression that had been predicted from the literature.

The difference between wild-type CD4⁺ T cells and *Jnk1*^{-/-}*Jnk2*^{-/-} or *dnJNK1*⁺*Jnk2*^{-/-} CD4⁺ T cells might reflect a difference in the relative abundance of sub-populations of T cells in these mice. Analysis by flow cytometry demonstrated that the peripheral T-cell populations of naive versus activated/memory cells from the *Rag1*^{-/-}*Jnk1*^{-/-}*Jnk2*^{-/-} or *dnJNK1*⁺*Jnk2*^{-/-} mice were comparable to those of the wild type. Furthermore, naive CD4⁺ T cells isolated from the mutant mice by CD45RB⁺CD44^{lo} or CD62L⁺ criteria still produced at least twofold more IL-2 than wild-type naive CD4⁺ T cells (data not shown). These data indicate that naive CD4⁺ T cells lacking JNK1 and JNK2 are intrinsically hyperresponsive to T-cell receptor (TCR) signalling during IL-2 expression. To test the degree of responsiveness of the JNK1- and JNK2-deficient cells, we stimulated wild-type or *dnJNK1*⁺*Jnk2*^{-/-} T cells with different doses of anti-CD3, with or without anti-CD28. The JNK-deficient T cells were not defective in IL-2 production even when treated with low-dose TCR or stimulated by CD28, and they seemed hyperresponsive to antigen receptor and CD28 co-stimulatory receptor ligation in cytokine secretion (Fig. 1c). These results indicate that JNK1 and JNK2 are not essential for IL-2 expression.

Originally identified as a T-cell growth factor, IL-2 mediates clonal expansion of T cells. In the absence of IL-2, T cells proliferate poorly¹⁰. Consistent with the IL-2 data, there was an enhanced

proliferation of *Jnk1*^{-/-}*Jnk2*^{-/-} or *dnJNK1*⁺*Jnk2*^{-/-} cells (Fig. 3a, and data not shown). IL-2 signalling allows cells to enter the cell cycle, in part by downregulating the cyclin-dependent kinase inhibitor p27 (ref. 11). Consistent with the enhanced IL-2 production and proliferation, 24 h after treatment with anti-CD3, with or without anti-CD28, *dnJNK1*⁺*Jnk2*^{-/-} CD4⁺ T cells expressed less p27 (Fig. 3b). IL-2, through its receptors, also induces multiple factors involved in cell growth and the inhibition of apoptosis, such as c-Myc and Bcl-2 (ref. 12). The *dnJNK1*⁺*Jnk2*^{-/-} T cells produced more Bcl-2 in response to anti-CD3, and more c-Myc in response to anti-CD3 plus anti-CD28, than the wild type (Fig. 3b). These data support our observation that T-cell activation was not defective in the absence of JNK1 and JNK2 and provide a potential mechanism where JNK is inhibitory to T-cell proliferation.

Our analysis appears to contradict an earlier observation that T cells deficient in a JNK kinase, MKK4, had defective IL-2 production in response to anti-CD3 and CD28 (ref. 13). However, it was also reported that JNK was activated normally in the absence of MKK4 (refs 13, 14). We identified the MAP kinase kinase, MKK7, which functions as a specific activator of JNK, unlike MKK4 (ref. 15). We proposed that MKK7 activates JNK in T cells to negatively regulate IL-2 induction. To assess this, we disrupted the *Mkk7* gene in ES cells. The *Mkk7*^{-/-} mice that we obtained by breeding the heterozygous *Mkk7*^{+/-} mice obtained from these ES cells die during embryogenesis (to be described elsewhere). Therefore, to analyse the function of T cells lacking MKK7 expression, we generated *Mkk7*^{-/-} ES cells, which lack MKK7 but retain MKK4 expression (Fig. 4a). In response to ultraviolet radiation and other stress signals, JNK activity was greatly reduced in the absence of MKK7, whereas p38 activity was not (Fig. 4b). To study MKK7-deficient lymphocytes, we complemented C57BL/6 or *Rag1*-deficient blastocysts with *Mkk7*^{-/-} ES cells and generated chimaeric mice. Similar to the *Jnk1*^{-/-}*Jnk2*^{-/-} mutants, B- and T-cell differentiation was unperturbed in *Rag1*^{-/-}*Mkk7*^{-/-} mice (data not shown). MKK7 appears to be important for JNK activation in T cells as only low levels of JNK activity were detected in *Mkk7*^{-/-} peripheral CD4⁺ T cells after treatment with anti-CD3 and anti-CD28 (data not

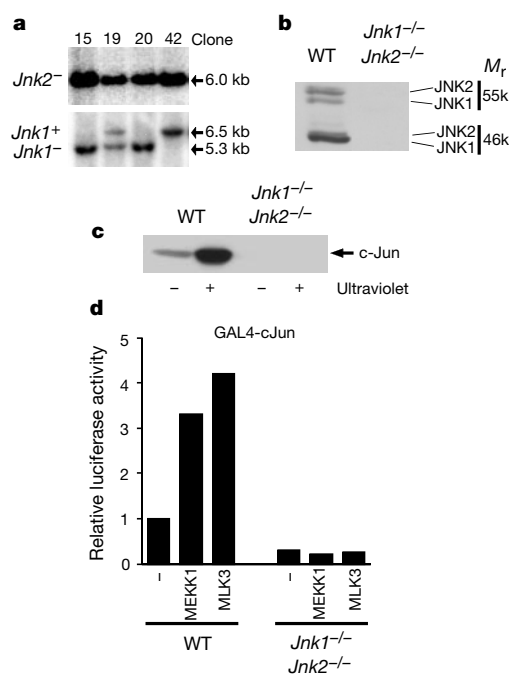


Figure 2 Identification and characterization of ES cells with disrupted *Jnk1* and *Jnk2* alleles. **a**, Blastocysts isolated from *Jnk1*^{+/-}*Jnk2*^{-/-} interbreeding were cultured with primary mouse embryonic fibroblasts in the presence of leukaemia inhibitory factor. ES cells were expanded and cloned. Genomic DNA was examined by Southern blot of the *Jnk1* and *Jnk2* alleles^{5,6}. As predicted, all the clones were *Jnk2*^{-/-} and individual clones corresponding to the three possible *Jnk1* genotypes were identified. **b**, The loss of JNK1 and JNK2 expression in *Jnk1*^{-/-}*Jnk2*^{-/-} clones was confirmed by immunoblot. **c**, The effect of exposure of wild-type (WT) and *Jnk1*^{-/-}*Jnk2*^{-/-} ES cells to ultraviolet-C radiation (60 J m⁻², followed by 1 h incubation) on JNK activity was examined using the substrate GST-c-Jun. **d**, Luciferase reporter gene expression was examined in transfection assays using GAL4-c-Jun. The effect of MEK1 and MLK3 is shown.

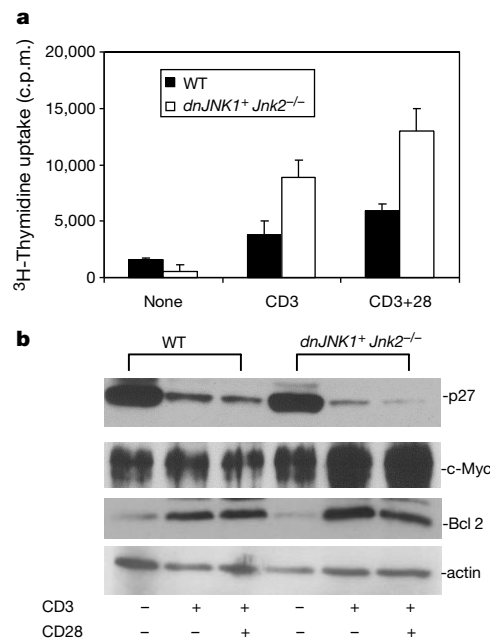


Figure 3 Enhanced proliferation by *dnJNK1*⁺*Jnk2*^{-/-} cells. **a**, CD4⁺ T cells isolated from wild-type (WT) or *dnJNK1*⁺*Jnk2*^{-/-} mice were treated with plate-bound anti-CD3 or anti-CD3 plus CD28 for three days. ³H-thymidine incorporation during the last 8 h was measured. **b**, CD4⁺ T cells were treated for 24 h and the amount of p27, c-Myc, Bcl-2 and actin was examined by immunoblot using antibodies purchased from Santa Cruz.

shown). Importantly, like *Jnk1*^{-/-}*Jnk2*^{-/-} T cells, activated *Mkk7*^{-/-} T cells were responsive to TCR stimulation in the presence or absence of CD28 ligation, produced more IL-2 and exhibited enhanced proliferation (Fig. 4c and d). These data indicate that MKK7 is indispensable for normal JNK activation in T cells mediated by the TCR and CD28 co-receptor, and that the MKK7–JNK signalling pathway is not required for T-cell activation and IL-2 production.

Consistent with this conclusion, we have reported that JNK activity is low in naive CD4⁺ T cells because the amount of JNK activator mRNA (MKK4 and MKK7) and the amount of JNK mRNA and protein extremely low⁸. Upon T-cell activation, the level of mRNA and protein, and the activity of these kinases are induced, peaking at about 48 h after TCR and CD28 co-receptor stimulation. In contrast, IL-2 production peaks during the first 24 h, a time when JNK, MKK4 and MKK7 expression is low. The amount of JNK protein and activity begins to increase following T-cell

activation, peaking at 48 h, a time when IL-2 expression decreases. Thus, the appearance of JNK signalling components in helper T cells is the result rather than the cause of T-cell activation, and indicates that this pathway may be important in the late phase of T-cell responses, that is, the differentiation of these cells into Th1 or Th2 effector subsets.

To test this hypothesis, we differentiated naive wild-type or *dnJNK1*⁺*Jnk2*^{-/-} helper T cells with anti-CD3 crosslinking in the presence of wild-type antigen-presenting cells (APC) and exogenous IL-2. Four days later, the outcome of the differentiation was assayed by restimulating differentiated cells with anti-CD3 and measuring the cytokines they produced. The wild-type effector cells, owing to the C57BL/6 × 129 F2 genetic background, produced mainly the Th1 effector cytokine IFN γ , indicating that they had preferentially developed along the Th1 lineage (Fig. 5a). On the other hand, while the JNK-deficient cells produced IFN γ , they

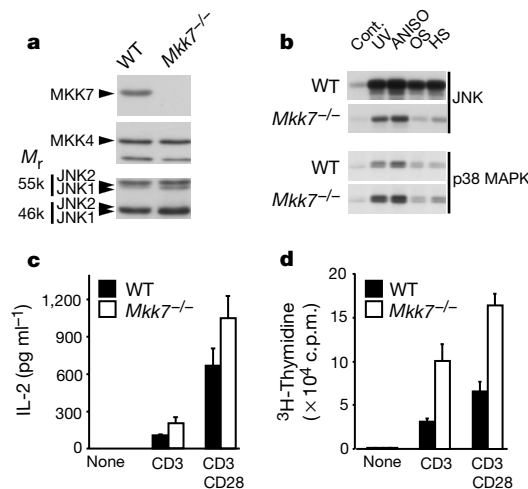


Figure 4 Increased IL-2 production and proliferation by *Mkk7*^{-/-} Th cells. **a**, The amount of MKK4, MKK7 and JNK expression in wild-type and *Mkk7*^{-/-} ES cells was examined by protein immunoblot analysis. **b**, Wild-type and *Mkk7*^{-/-} ES cells were exposed to 60 J m⁻² ultraviolet radiation, 1 μ g ml⁻¹ anisomycin (ANISO), 300 mM sorbitol (OS) or 42 °C (HS). After incubation for 1 h, JNK and p38 MAP kinase activity was measured in an immunocomplex kinase using GST–cJun and GST–ATF2 as substrates, respectively.

c,d, *Mkk7*^{-/-} CD4⁺ T cells were purified from C57BL/6–*Mkk7*^{-/-} chimaera mice using the Ly9.1 allotypic marker. The cells were cultured and incubated without or with plate-bound anti-CD3 or anti-CD3 plus CD28 antibodies. The production of IL-2 was measured by ELISA (**c**) and proliferation was monitored by measurement of ³H-thymidine incorporation (**d**).

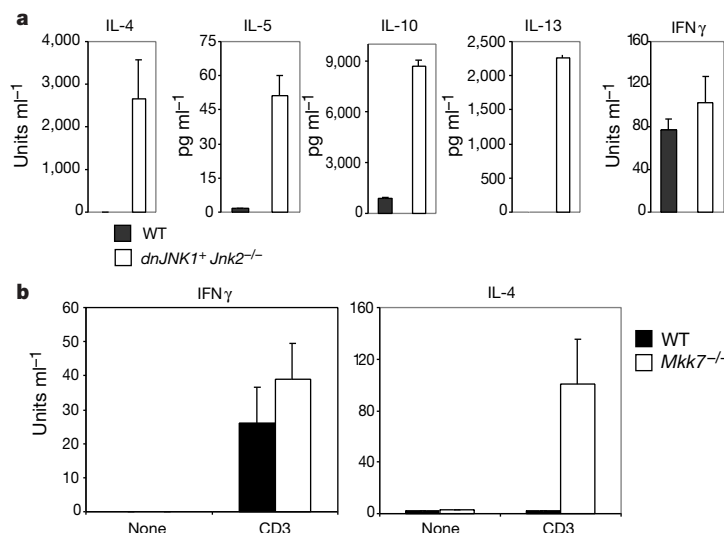


Figure 5 Defective Th cell differentiation in the absence of JNK signalling. **a**, Naive wild-type (WT) or *dnJNK1*⁺*Jnk2*^{-/-} Th cells were differentiated *in vitro* with anti-CD3, irradiated wild-type APC and IL-2 for four days. The differentiated cells were restimulated with anti-

CD3 and cytokine production measured by ELISA. **b**, Wild-type or *Mkk7*^{-/-} CD4⁺ T cells were treated with anti-CD3 for four days and cytokine production was measured by ELISA.

developed a pronounced Th2 response, producing large amounts of all the Th2 cytokines that we measured, including IL-4, 5, 10 and 13 (Fig. 5a). Such a Th2 response is probably driven by IL-4 and IL-13 production from the knockout T cells in the course of differentiation in response to TCR activation (data not shown). In agreement with the above observation, we found that after anti-CD3 activation, *Mkk7*^{-/-} CD4 T cells, while producing similar amounts of IFN γ to the wild-type cells, also secreted markedly increased amounts of IL-4 (Fig. 5b). These data indicate that the JNK signalling pathway may function as a negative regulator of Th2 responses. Although not required for the initial T-cell activation, JNK is required for polarized differentiation of Th cells into the Th1 lineage, in keeping with our analysis of JNK1-deficient animals⁵.

We initiated these studies with the expectation, based on other studies^{3,4}, that the JNK pathway was a positive regulator of IL-2 production. Unexpectedly, we found both in transfection studies and in three animal models in which T cells from mice lack JNK protein and enzyme activity that production of IL-2 was not reduced. Indeed more IL-2 was produced and T cells proliferated more vigorously than wild-type T cells. However, lack of JNK affects Th cell differentiation. While JNK2 has a function in IFN γ production by Th1 effector cells in response to IL-12 (ref. 6), T cells deficient in JNK1 or both JNK1 and JNK2 produced more Th2 cytokines, indicating that they are essential regulators for polarized Th1 differentiation. This, and the demonstration that the amount of JNK protein and activity increases after T-cell activation, indicates a new model for the role of JNK in T-cell immune responses, that is, the regulation of Th cell differentiation. In support of this, we have identified a new substrate for JNK1, NFATc1, which serves as an indispensable activator for Th2 cell differentiation¹⁶. We suggest that the role of JNK is to potentiate polarized T-cell differentiation into the Th1 lineage by inhibiting Th2 cytokine production. Consistent with this, the main phenotypes of *Jnk1*^{-/-}, *Jnk2*^{-/-}, *Mkk7*^{-/-} and *dnJNK1*⁺*Jnk2*^{-/-} T cells are abnormalities in effector cytokine expression. Thus, JNK is required for effector T-cell function but not for T-cell activation. □

Methods

Helper T-cell activation and differentiation

The methods used for isolation and treatment of peripheral CD4⁺ T cells, the measurement of IL-2 production by ELISA (Pharmingen), the measurement of IL-2 mRNA by polymerase chain reaction with reverse transcription (RT-PCR), and ³H-thymidine incorporation have been described⁵. To differentiate helper T cells *in vitro*, CD4 T cells from wild-type or *dnJNK1*⁺*Jnk2*^{-/-} mice were enriched, and the CD62L⁺ naive population was further purified by magnetic assisted cell sorting. These cells (1 × 10⁶ ml⁻¹) were incubated with immobilized anti-CD3, irradiated T-depleted wild-type splenocytes as APC (1 × 10⁶ ml⁻¹) and IL-2 (30 units ml⁻¹) for four days. The differentiated cells were then extensively washed, counted and restimulated with anti-CD3. Cytokine production was measured by ELISA.

Generation of *Mkk7*^{-/-} ES cells and isolation of *Mkk7*^{-/-} T cells

The *Mkk7* targeting construct replacing three exons with a ploxP-flanked puromycin-resistance gene driven by the PGK promoter was transfected into W9.5 ES cells and homologous recombinants were identified by Southern blot. After removal of the ploxP-flanked puromycin gene in the targeted locus with a *cre* expression vector (Gibco BRL), the second *Mkk7* allele was then targeted with the same replacement vector and *Mkk7*^{-/-} ES cells were identified by Southern blot. *Mkk7*^{-/-} ES cells were used to reconstitute C57BL/6 or *Rag1*-deficient blastocysts. The chimaeric mice were identified by coat colour. *Mkk7*^{-/-} peripheral T cells in B6 chimaeras were isolated using the Ly9.1 allotypic marker.

Received 28 January; accepted 22 March 2000.

1. Ip, Y. T. & Davis, R. J. Signal transduction by the c-Jun N-terminal kinase (JNK)—from inflammation to development. *Curr. Opin. Cell Biol.* **10**, 205–219 (1998).
2. Li, B., Tournier, C., Davis, R. J. & Flavell, R. A. Regulation of IL-4 expression by the transcription factor JunB during T helper cell differentiation. *EMBO J.* **18**, 420–432 (1999).
3. Su, B. *et al.* JNK is involved in signal integration during costimulation of T lymphocytes. *Cell* **77**, 727–736 (1994).
4. Matsuda, S., Moriguchi, T., Koyasu, S. & Nishida, E. T-Lymphocyte activation signals for Interleukin-2 production involve activation of MKK6-p38 and MKK7-SAPK/JNK signaling pathways sensitive to cyclosporin A. *J. Biol. Chem.* **273**, 12378–123782 (1998).
5. Dong, C. *et al.* Defective T-cell differentiation in the absence of Jnk1. *Science* **282**, 2092–2095 (1998).
6. Yang, D. D. *et al.* Differentiation of CD4⁺ T cells to Th1 cells requires MAP kinase JNK2. *Immunity* **9**, 575–585 (1998).

7. Kuan, C. Y. *et al.* The Jnk1 and Jnk2 protein kinases are required for regional specific apoptosis during early brain development. *Neuron* **22**, 667–676 (1999).
8. Weiss, L. *et al.* Regulation of c-Jun NH₂-terminal kinase (JNK) gene expression during T-cell activation. *J. Exp. Med.* **191**, 139–145 (2000).
9. Rincon, M. *et al.* The JNK pathway regulates the *in vivo* deletion of immature CD4(+)CD8(+) thymocytes. *J. Exp. Med.* **188**, 1817–1830 (1998).
10. Schorle, H., Holschke, T., Hunig, T., Schimpl, A. & Horak, I. Development and function of T cells in mice rendered interleukin-2 deficient by gene targeting. *Nature* **352**, 621–624 (1991).
11. Nourse, J. *et al.* Interleukin-2-mediated elimination of the p27Kip1 cyclin-dependent kinase inhibitor prevented by rapamycin. *Nature* **372**, 570–573 (1994).
12. Gomez, J., Gonzalez, A., Martinez A., C. & Rebollo, A. IL-2-induced cellular events. *Crit. Rev. Immunol.* **18**, 185–220 (1998).
13. Nishina, H. *et al.* Impaired CD28-mediated interleukin 2 production and proliferation in stress kinase SAPK/ERK1 kinase (SEK1)/mitogen-activated protein kinase kinase 4 (MKK4)-deficient T lymphocytes. *J. Exp. Med.* **186**, 941–953 (1997).
14. Swat, W. *et al.* SEK1/MKK4 is required for maintenance of a normal peripheral lymphoid compartment but not for lymphocyte development. *Immunity* **8**, 625–634 (1998).
15. Tournier, C., Whitmarsh, A. J., Cavanagh, J., Barrett, T. & Davis, R. J. Mitogen-activated protein kinase kinase 7 is an activator of the c-Jun NH₂-terminal kinase. *Proc. Natl Acad. Sci. USA* **94**, 7337–7342 (1997).
16. Chow, C. -W., Dong, C., Flavell, R. A., & Davis, R. J. JNK inhibits targeting of protein phosphatase calcineurin to NFATc1. *Mol. Cell. Biol.* (in the press).

Acknowledgements

The authors thank D. Y. Loh and C. L. Stewart for providing reagents; L. Evangelisti, D. Butkus, C. Hughes and J. Stein for technical assistance; and F. Manzo for secretarial work. Supported by NIH grants and the Howard Hughes Medical Institute. C.D. is a recipient of an Arthritis Foundation Postdoctoral Fellowship. C.D. and D.D.Y. were Associates, C.T. is an Associate, and R.J.D. and R.A.F. are Investigators of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to R.A.F.

A role for excreted quinones in extracellular electron transfer

Dianne K. Newman* & Roberto Kolter

Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA

Respiratory processes in bacteria are remarkable because of their ability to use a variety of compounds, including insoluble minerals, as terminal electron acceptors¹. Although much is known about microbial electron transport to soluble electron acceptors, little is understood about electron transport to insoluble compounds such as ferric oxides^{2,3}. In anaerobic environments, humic substances can serve as electron acceptors and also as electron shuttles to ferric oxides^{4–6}. To explore this process, we identified mutants in *Shewanella putrefaciens* that are unable to respire on humic substances. Here we show that these mutants contain disruptions in a gene that is involved in the biosynthesis of menaquinone. During growth, the wild type releases a menaquinone-related redox-active small molecule into the medium that complements the mutants. This finding raises the possibility that electron transfer to a variety of oxidants, including poorly soluble minerals, may be mediated by microbially excreted quinones that have yet to be identified.

Because of its importance in a multitude of environmental processes, the mechanism of electron transfer by microbes to poorly soluble minerals for anaerobic respiration has been a subject of intense study^{7,8}. Various proteins involved in this process, including outer membrane proteins^{9,10} and periplasmic and extracellular cytochromes^{11,12}, have been identified in organisms that respire by

* Present address: Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA.

using ferric and/or manganese oxides as terminal electron acceptors. Nevertheless, the details of how these proteins function remain enigmatic and are still the subject of debate. For example, the provocative hypothesis that extracellular cytochromes can serve as ferric reductases¹² has been questioned recently, in part, because of concerns that this would be energetically too costly¹³. But the fact that exogenous extracellular molecules (such as humic acids) can participate in electron transfer to insoluble minerals indicates that they may make a significant contribution to mineral dissolution in many environments. Whether microbially produced extracellular molecules have a similar role remains an important question.

To study humic acid reduction at the molecular level, we chose *S. putrefaciens* strain MR-1 as a model organism, as its genome has been sequenced and another member of the *Shewanella* genus has been shown to respire humate⁴. As quinones are the electron accepting moieties of humic substances¹⁴, we screened 7,000 transposon-generated mutants of *S. putrefaciens* strain MR-1 (ref. 15) for their ability to use the humic acid analogue, anthraquinone-2,6-disulphonate (AQDS)¹⁶ as an electron acceptor during anaerobic growth on lactate. Because the colour of AQDS changes from clear to orange when it becomes reduced, mutants were identified as cultures that remained clear or were slow to turn orange (Fig. 1a). Out of 7,000 mutants, we identified 2 (H1 and H2) that were completely defective in AQDS reduction. Both mutants were also severely hampered in their ability to reduce humate (Aldrich), providing genetic evidence of a common biochemical basis for AQDS and humate reduction (Fig. 1b). Each of the two mutants contained an insertion in a different site of a homologue of the *Escherichia coli* *menC* gene, that encodes *o*-succinylbenzoic acid

(OSB) synthase¹⁷. This enzyme is involved in the synthesis of menaquinone, a lipophilic naphthoquinone with a prenyl side chain of variable length that participates in the transfer of electrons to low-potential electron acceptors (Fig. 2a).

Without *menC*, we expected that the mutants would be unable to produce menaquinone and would therefore be unable to grow on a variety of low-potential electron acceptors¹⁸. Lipid extractions of cultures grown aerobically showed that this was the case: both ubiquinone and menaquinone were found in the wild type (as assayed by thin layer chromatography (TLC)), but only ubiquinone was found in the mutants (Fig. 2b). Also, when the mutants were cultured on lactate minimal medium, growth only occurred in the presence of oxygen or nitrate (although growth on nitrate was down relative to the wild type), but not in the presence of manganese oxide (MnO₂), fumarate, thiosulphate, sulphite, dimethyl sulphoxide (DMSO) or ferrihydrite (Fe(OH)₃). The addition of menaquinone or 1,4-dihydroxy-2-naphthoic acid (DHNA, a biosynthetic precursor of menaquinone downstream of the predicted block in the pathway) restored the ability of the mutants to grow on AQDS. These results show that menaquinone is used in the electron transport chain of *S. putrefaciens* to pass electrons to a terminal reductase that reduces AQDS and humic acid.

Given previous speculation on the role of extracellular cytochromes in electron transport to insoluble minerals¹², we proposed that wild-type *S. putrefaciens* might produce an extracellular quinone that could transfer electrons to oxidants outside the cell. We therefore performed a complementation test in which we streaked the wild type and the mutants on a plate containing AQDS. When the mutants were streaked in isolation they did not grow and the

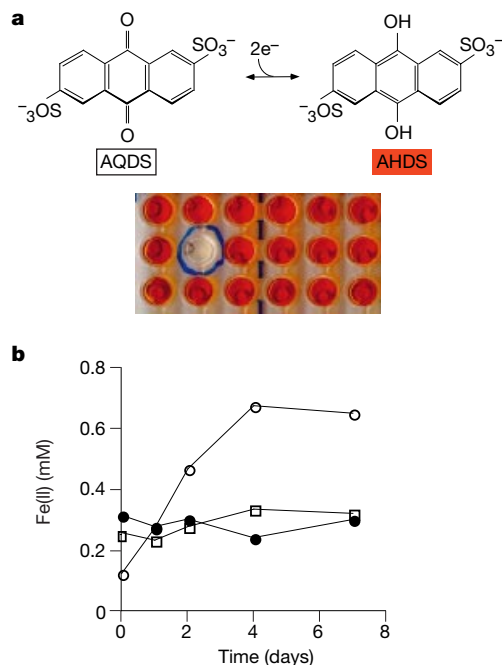


Figure 1 Identification of AQDS/humate-reduction mutants. **a**, Genetic screen used to isolate mutants defective in AQDS reduction. Because AQDS is clear when oxidized and orange when reduced to AHDS, mutants were identified as those which remained clear. **b**, The AQDS-reduction defective mutants were tested for their ability to grow on and reduce humic substances as measured indirectly by the release of Fe(II) after reacting the reduced humate with Fe(III). The wild-type (open circles) rapidly reduced humate, whereas the mutant (open squares) did not. Data shown are representative of triplicate cultures, with filled circles showing an uninoculated control.

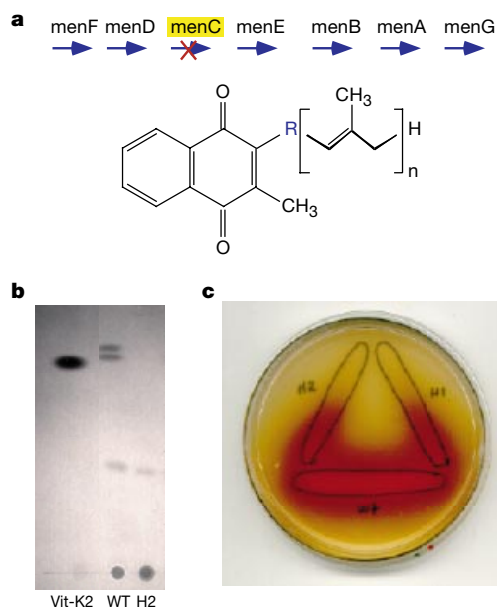


Figure 2 Mutants defective in menaquinone biosynthesis are rescued by a diffusible compound. **a**, The biosynthesis of menaquinone in *E. coli* is controlled by seven genes. The mutants defective in AQDS reduction contained independent transposon insertions in a homologue to *menC*, an intermediate along this pathway. The structure of menaquinone is shown. **b**, Thin layer chromatography. Lane 1, a menaquinone control (vitamin K2). Lane 2, spots from extracts of the wild-type (WT): the band half-way up the plate is ubiquinone, the spots at the top are menaquinones. Lane 3, a single spot of ubiquinone from one of the *menC* mutants (H2), but menaquinones are not produced. **c**, Anaerobic growth of wild-type *Shewanella putrefaciens* (WT) and two *menC* mutants (H1, H2) on lactate with AQDS as the electron acceptor. Reduction of AQDS, as determined by the bright orange colour, requires a diffusible molecule produced by the wild-type that complements the mutants.

plate remained clear; however, when streaked in close proximity to the wild type the mutants grew and were able to reduce AQDS to AHDS (Fig. 2c). This result implies that they were complemented by an extracellular diffusible factor, and this was confirmed in liquid culture by measuring humic acid reduction by the mutants in the presence of spent wild-type supernatants (data not shown).

To characterize the diffusible factor produced by wild-type *S. putrefaciens*, we used a bioassay (complementation of H1 or H2 on an anaerobic AQDS plate). First, we tested whether other organisms could produce the diffusible factor. Complementation occurred with other *Shewanella* species (for example, *Shewanella alga* or a newly isolated *Shewanella* strain from Woods Hole, Massachusetts) and with *Vibrio cholerae*, but not with *E. coli* or *Pseudomonas fluorescens*. In addition, no complementation was seen when a *menC* mutant of *E. coli* was placed in proximity to wild-type *E. coli* and growth was assayed with fumarate as the electron acceptor. This suggests that, under these conditions, a diffusible factor that complements *menC* mutants is produced by some, but not all, of the gamma Proteobacteria.

We found that supernatants from wild-type *S. putrefaciens* cultures (growing in exponential phase under either aerobic or anaerobic conditions) complemented the mutants, whereas supernatants from the *Shewanella menC* mutants (growing under aerobic conditions) did not. When wild-type *S. putrefaciens* supernatants were treated with various proteases, they remained active, showing that the diffusible factor is not a protein. Furthermore, the diffusible factor was found to be smaller than relative molecular mass 10,000 (M_r 10K), as activity was retained in the supernatants after ultra-

filtration. Thin layer chromatography revealed that menaquinone was not present in the supernatants of wild-type *S. putrefaciens* (Fig. 3a), and, therefore, that it is not the diffusible factor. Menaquinone is a highly hydrophobic compound that is normally found within the lipid bilayer. We detected a large active spot (determined by its ability to complement H2) that remained at the bottom of the TLC plate in the same position as DHNA (Fig. 3a), but further TLC analysis showed that the active spot was distinct from DHNA (Fig. 3b). Thus, the product of the *menC* gene appears to be required for the synthesis of both the diffusible factor and menaquinone.

Because *menC* encodes an enzyme that participates in the biosynthesis of menaquinone, we proposed that the diffusible factor might be a quinone. Indeed, the TLC active spots (Fig. 3a and b) were identified by absorption under ultraviolet light (254 nm), a standard method for detection of quinones¹⁹. High performance liquid chromatography (HPLC) separation of the TLC active spot and the concentrated supernatant yielded a fraction with high ultraviolet absorbance (Fig. 3c, insert). This fraction was highly active in the complementation assay and also turned orange upon reduction with hydrogen gas in the presence of palladium, as measured by absorption at 350 nm (Fig. 3c, bar graph). The colour change was not seen in the presence of only palladium or hydrogen. Such a colour change is consistent with the presence of a redox-active quinone⁴. Fractions collected on either side of the absorption peak yielded both negligible complementation (Fig. 3c, circles) and colour change. Mass spectrometry data on the active TLC spots and HPLC fraction revealed that the main constituent has a molecular mass in the range expected for a small quinone (M_r 150–300).

In conclusion, we have shown that there is a common biochemical basis for AQDS and humic acid respiration involving menaquinone. Furthermore, a non-proteinaceous small compound that has characteristics similar to a quinone and can be excreted into the medium is involved in electron transfer to AQDS and humic acid. Possibly this compound may simply enable the cells to synthesize menaquinone, but it is also possible that it is directly involved in extracellular electron transfer. In the latter case, small microbially produced quinones may represent the elusive extracellular electron shuttles that allow microbes to respire minerals. □

Methods

Identification of mutants

Transposon insertions were made by mating *S. putrefaciens* strain MR-1 with *E. coli* strain β -2155, containing pBSL180, a suicide plasmid carrying a mini-Tn10 transposon derivative marked with a kanamycin (Km) cassette²⁰. Exconjugants were selected on lactate minimal medium²¹ supplemented with 50 $\mu\text{g ml}^{-1}$ Km under aerobic conditions. The mutants were inoculated into 96-well microtitre plates filled with LB medium (+ Km) and allowed to grow overnight. Three microlitres of each culture was transferred into 96-well microtitre plates containing lactate minimal medium supplemented with AQDS and screened for mutants defective in AQDS reduction. Strains that remained clear under these conditions (after triple checks) were assayed for their ability to reduce humic acid. A polymerase chain reaction (PCR)-based technique²² was used to identify the sequences flanking the transposon insertions of two strains that were completely unable to reduce AQDS and humic acid (H1 and H2). These sequences were submitted to a BLAST²³ search of the *Shewanella putrefaciens* genome on the TIGR web site (<http://www.tigr.org>) and the precise site of the transposon insertion was identified. Contigs that contained matches were then analysed with the NCBI program ORF-Finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>). Two independent PCR reactions and subsequent sequence analyses confirmed that the transposon insertions in H1 and H2 lay in a homologue to the *menC* gene (40% identity, 55% positives, amino-acid pairwise alignment using the NCBI program BLAST2).

Chemical analyses

Humate reduction was assayed indirectly by measuring the amount of ferric iron reduced in the presence of the humic substances^{4,24}. Quinones were assayed by TLC by adapting a described protocol¹⁹. Silica Gel 60 F₂₅₄ glass-backed TLC plates were spotted with the samples of interest and eluted with an organic solvent (Fig. 3a, petroleum ether:diethyl ether, 85:15; Fig. 3b, ethanol:diethyl ether, 2:1). Menaquinones were identified by brief irradiation with ultraviolet light (254 nm) as purple spots on a green fluorescent background that migrated to an R_f value of ~ 0.7 , and ubiquinones were identified as spots that migrated to an R_f value of ~ 0.3 . MnO_2 and $\text{Fe}(\text{OH})_3$ were synthesized as described²⁵. Activity tests were performed by scraping the spots off the plates and eluting the

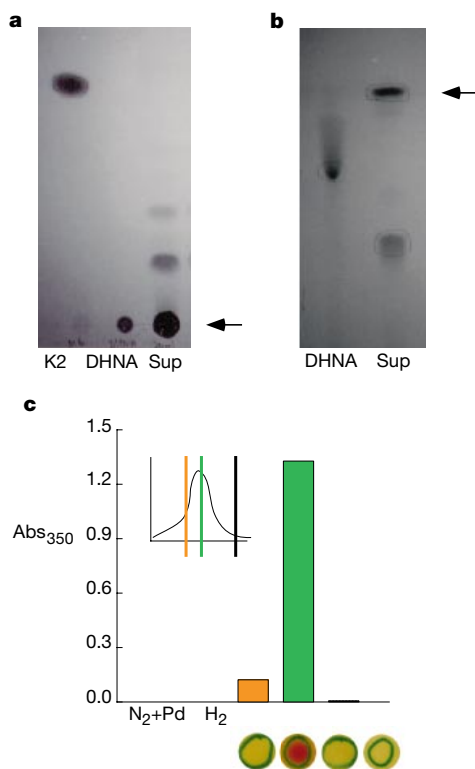


Figure 3 Nature of the diffusible compound. **a**, TLC of supernatants from exponentially growing wild-type cultures. No menaquinone was measured in the supernatants. The active spot (denoted by arrow) remained at the bottom of the plate next to DHNA. **b**, TLC of supernatants from exponentially-growing wild-type cultures. The active spot is distinct from DHNA. **c**, Redox sensitivity of the diffusible compound. Inset, chromatogram showing the relative positions of fractions that were sampled for redox analysis (Abs_{350}). Circles below the bar graph show the relative ability of these fractions to complement the *menC* mutants in the bioassay. The circle on the right is a negative control (no addition).

compounds with 100% methanol. The eluant was then concentrated with a speed-vac and used in the complementation assay with H₂. Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) and LC/MS instruments were used for mass spectrometry of active samples.

Complementation assays

Complementation assays were performed by spotting 10 µl of an overnight culture of H₂ (grown aerobically) on a lactate minimal medium plate (amended with 5 mM AQDS) plus 10 µl of the appropriate compound (for example, DHNA resuspended in 100% methanol, Vitamin K₂ resuspended in dimethylformamide, added to a final concentration of 1–10 µM). TLC or HPLC fractions were concentrated under a speed-vac and resuspended in 100% methanol before their addition. When multiple strains were used in complementation assays, ~3 µl of an overnight culture was painted onto circumscribed areas of the plates using sterile loops. The plates were dried in a laminar flow hood and incubated overnight under anaerobic conditions. Growth and AQDS reduction were scored the next day.

Redox determination

Ten litres of supernatants from wild-type *S. putrefaciens* cultures were concentrated by passage over a low pressure C₁₈ column and eluted with 50% methanol. This concentrated fraction was then extracted with ether to remove highly hydrophobic compounds from the aqueous phase. The aqueous phase was then run over a large preparative C₁₈ column on a Waters HPLC and eluted with a linear gradient of 0–100% methanol in water. Absorption at 214 nm and 254 nm was recorded and 10 ml fractions were collected at a flow rate of 10 ml min⁻¹. Selected fractions were checked for activity using the bioassay. Three fractions were chosen for subsequent analysis: before, during, and after the peak of activity. These fractions were dried down and resuspended in 2 ml anoxic bicarbonate buffer (0.25%). About 0.5 g of palladium shot (Aldrich) was added to 1 ml of these solutions; they were flushed with H₂ for 5 min and capped in small anaerobic vials. A Beckman DU-640 spectrophotometer was used to measure the change in absorption of these solutions at 350 nm.

Received 14 December 1999; accepted 10 February 2000.

1. Brock, T. D. & Madigan, M. T. *Biology of Microorganisms* 6th edn (Prentice Hall, New Jersey, 1991).
2. Gennis, R. B. & Stewart, V. in *Escherichia coli and Salmonella: Cellular and Molecular Biology* vol 1 (ed. Neidhardt, F. C.) 217–261 (ASM, Washington DC, 1996).
3. Lovley, D. R. Dissimilatory Fe(III) and Mn(IV) reduction. *Microbiol. Rev.* **55**, 259–287 (1991).
4. Lovley, D. R., Coates, J. D., Blunt-Harris, E. L., Phillips, E. J. P. & Woodward, J. C. Humic substances as electron acceptors for microbial respiration. *Nature* **382**, 445–448 (1996).
5. Coates, J. D. *et al.* Recovery of humic-reducing bacteria from a diversity of environments. *Appl. Environ. Microbiol.* **64**, 1504–1509 (1998).
6. Benz, M., Schink, B. & Brune, A. Humic acid reduction by *Propionibacterium freudenreichii* and other fermenting bacteria. *Appl. Environ. Microbiol.* **64**, 4507–4512 (1998).
7. Caccavo, F. Jr. *et al.* Role of hydrophobicity in adhesion of the dissimilatory Fe (III)-reducing bacterium *Shewanella alga* to amorphous Fe (III) oxide. *Appl. Environ. Microbiol.* **63**, 3837–3843.
8. Arnold, R., DiChristina, T. J. & Hoffmann, M. R. Reductive dissolution of Fe (III) oxides by *Pseudomonas* sp. 200. *Biotechnol. Bioeng.* **32**, 1081–1096 (1988).
9. Gaspard, S., Vazquez, F. & Holliger, C. Localization and solubilization of the iron(III) reductase of *Geobacter sulfurreducens*. *J. Bacteriol.* **64**, 3188–3194.
10. Beliaev, A. S. & Saffarini, D. A. *Shewanella putrefaciens* *mtrB* encodes and outer membrane protein required for Fe(III) and Mn(IV) reduction. *J. Bacteriol.* **180**, 6292–6297 (1998).
11. Myers, C. R. & Myers, J. M. Localization of cytochromes to the outer membrane of anaerobically grown *Shewanella putrefaciens* MR-1. *J. Bacteriol.* **174**, 3429–3438 (1992).
12. Seeliger, S., Cord-Ruwisch, R. & Schink, B. A periplasmic and extracellular c-type cytochrome of *Geobacter sulfurreducens* acts as a ferric iron reductase and as an electron carrier to other acceptors or to partner bacteria. *J. Bacteriol.* **180**, 3686–3691 (1998).
13. Lloyd, J. R., Blunt-Harris, E. L. & Lovley, D. R. The periplasmic 9.6 kDa c-type cytochrome is not an electron shuttle to Fe(III). *J. Bacteriol.* **181**, 7647–7649 (1999).
14. Scott, D. T., McKnight, D. M., Blunt-Harris, E. L., Kolesar, S. E. & Lovley, D. R. Quinone moieties act as electron acceptors in the reduction of humic substances by humics-reducing microorganisms. *Environ. Sci. Technol.* **32**, 2984–2989 (1998).
15. Myers, C. R. & Nealson, K. H. Bacterial manganese reduction and growth with manganese oxide as the sole electron acceptor. *Science* **240**, 1319–1320 (1988).
16. Tratnyek, P. G. & Macalady, D. L. Abiotic reduction of nitro aromatic pesticides in anaerobic laboratory systems. *J. Agri. Food Chem.* **37**, 248–254 (1989).
17. Meganathan, R. in *Escherichia coli and Salmonella: Cellular and Molecular Biology* vol. 1 (ed. Neidhardt, F. C.) 642–656 (ASM, Washington DC, 1996).
18. Myers, C. R. & Myers, J. M. Role of menaquinone in the reduction of fumarate, nitrate, iron(III) and manganese(IV) by *Shewanella putrefaciens* MR-1. *FEMS Microbiol. Lett.* **114**, 215–222 (1993).
19. Collins, M. D. Analysis of isoprenoid quinones. *Methods Microbiol.* **18**, 329–366 (1985).
20. Alexeyev, M. F. & Shokolenko, I. N. Mini-Tn10 transposon derivatives for insertion mutagenesis and gene delivery into the chromosome of Gram-negative bacteria. *Gene* **160**, 59–62 (1995).
21. Myers, C. R. & Nealson, K. H. Respiration-linked proton translocation coupled to anaerobic reduction of manganese (IV) and iron (III) in *Shewanella putrefaciens* MR-1. *J. Bacteriol.* **172**, 6232–6238 (1990).
22. O'Toole, G. A. *et al.* Genetic approaches to the study of biofilms. *Methods Enzymol.* **310**, 91–109 (1999).
23. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
24. Stookey, L. L. Ferrozine: a new spectrophotometric reagent for iron. *Anal. Chem.* **42**, 779–781 (1970).
25. Lovley, D. R. & Phillips, E. J. Novel mode of microbial energy metabolism: organic carbon oxidation coupled to dissimilatory reduction of iron and manganese. *Appl. Environ. Microbiol.* **54**, 1472–480 (1988).

Acknowledgements

We thank C. Myers for providing *S. putrefaciens* strain MR-1 and members of the Kolter lab, P. Belshaw, F. Morel, B. Schink and D. Lovley for advice and encouragement. Preliminary sequence data was obtained from The Institute for Genomic Research website (<http://www.tigr.org>). Sequencing of *S. putrefaciens* was accomplished with support from the Department of Energy. This research was supported by the Office of Naval Research, NIH and a postdoctoral fellowship to D.K.N. by the Cancer Research Fund of the Damon Runyon Walter Winchell Foundation.

Correspondence and requests for materials should be addressed to D.K.N. (e-mail: dkn@gps.caltech.edu).

Structural insights into the stereochemistry of the cyclooxygenase reaction

James R. Kiefer*, Jennifer L. Pawlitz*, Kirby T. Moreland*, Roderick A. Stegeman*, William F. Hood*, James K. Gierse*, Anna M. Stevens*, Douglas C. Goodwin†, Scott W. Rowlinson†, Lawrence J. Marnett†, William C. Stallings* & Ravi G. Kurumbail*

* Searle Discovery Research, Monsanto Company, 700 Chesterfield Parkway North, St Louis, Missouri 63198, USA

† Department of Biochemistry, Vanderbilt School of Medicine, Vanderbilt University, Nashville, Tennessee 37232, USA

Cyclooxygenases are bifunctional enzymes that catalyse the first committed step in the synthesis of prostaglandins, thromboxanes and other eicosanoids^{1–3}. The two known cyclooxygenases isoforms share a high degree of amino-acid sequence similarity^{1–4}, structural topology^{5–7} and an identical catalytic mechanism^{1–3}. Cyclooxygenase enzymes catalyse two sequential reactions in spatially distinct, but mechanistically coupled active sites^{8–11}. The initial cyclooxygenase reaction converts arachidonic acid (which is achiral) to prostaglandin G₂ (which has five chiral centres). The subsequent peroxidase reaction reduces prostaglandin G₂ to prostaglandin H₂. Here we report the co-crystal structures of murine apo-cyclooxygenase-2 in complex with arachidonic acid and prostaglandin. These structures suggest the molecular basis for the stereospecificity of prostaglandin G₂ synthesis.

Cyclooxygenase (COX) enzymes require one catalytic turnover from their peroxidase active site to initiate the cyclooxygenase reaction, and His 207 (residue numbering based on sheep seminal vesicle COX-1) is critical for peroxidase activity¹². We made the His207Ala variant of COX-2 and co-crystallized it with arachidonic acid (AA), which allowed us to study substrate binding to the cyclooxygenase active site in its unaltered state. The crystal structure of this complex is refined at 2.4 Å resolution and shows few deviations from the previously reported structure of the COX-2 holoenzyme⁷. Superposition of the two structures over 552 Cα atoms results in an r.m.s. deviation of 0.47 Å. The differences between this and previous structures included a decreased volume of the membrane channel (helices B–D, residues 86–124) and side-chain rotamer changes in the two active sites.

The cyclooxygenase active site of the enzyme is located at the end of a hydrophobic channel in the protein that extends from the membrane-binding region towards the haem. Substrate is bound tightly in the cyclooxygenase active site, giving rise to well-defined electron density and average thermal parameters comparable to those of the protein. Distinct kinks in the electron density maps correspond to the four *cis* double bonds of the substrate (Fig. 1a). Inspection of the packing in the active site using the program

PROBE¹³ reveals that AA forms extensive van der Waals contacts with the protein and is virtually sequestered from solvent. Only the side chain of Leu 531 adopts a different rotamer conformation, accommodating the ω -end of the substrate. Differences in the sequence and local structure of adjacent residues 115–125 (relative to COX-1) facilitate the altered conformation of Leu 531 in COX-2.

Intriguingly, the observed orientation of the substrate is opposite to that of predictions made on the basis of biochemical and mutagenesis studies. The carboxylate end of the substrate is unambiguously bound and forms strong hydrogen bonds to the side chains of Tyr 385 and Ser 530 (Fig. 1b). The Arg 120 side chain forms van der Waals contacts with the ω -end of the substrate rather than the predicted ion pair with the carboxylate of AA. The observed substrate orientation, however, is not viable for catalysis. Spectroscopic and mutagenesis experiments have established that a catalytically essential radical forms on Tyr 385 (refs 14, 15). The reaction is initiated by abstraction of a hydrogen from C13 of AA by a protein radical^{18,16}. In the current structure, the phenolic oxygen of Tyr 385 is 10.5 Å from C13 of AA—too far for chemistry to occur. These studies and data discussed below suggest that the observed AA conformation most probably represents a non-productive binding mode of the substrate in the enzyme active site (see also Supplementary Information).

Thin solvent channels emanate from each of the cyclooxygenase active sites of the COX-2 dimer and converge to form a large, partially enclosed reservoir of solvent molecules at the dimer interface (Fig. 2). The membrane channel was proposed to act as the route into and out of the cyclooxygenase active site for substrate and product^{5–7}. If true, this would deposit prostaglandin G₂ (PGG₂) between the two membrane-binding regions of the COX-2 dimer. The solvent reservoir at the dimer interface could then function as a conduit for PGG₂, facilitating its direct transport to either peroxidase active site. Exogenous PGG₂ is unable to compete with PGG₂ generated by microsome-associated COX-1 for access to the peroxidase active site¹⁷; in contrast, exogenous PGG₂ is readily metabolized by detergent-solubilized COX-1. Consistent with our proposal, these observations emphasize the importance of mem-

brane association of the enzyme for efficient transduction of PGG₂ from one active site to another.

In the second experiment, we measured diffraction data to 3.0 Å resolution from a wild-type apoenzyme COX-2 crystal grown in the presence of AA. This structure also has AA bound in the orientation discussed above; however, additional features in the electron density can not be explained by the presence of AA alone, but are consistent with the chemical structures of PGG₂ or PGH₂ (Fig. 3a). Although no exogenous haem was added during the crystallization experiment, a 5–10% contaminant of haem-containing protein molecules in ‘apoenzyme’ preparations can be detected spectroscopically (J.K.G., unpublished data). Concentrations of AA in excess of 400 μ M cause dissociation of haem from COX-1 (ref. 18). Therefore, the diffusion of residual haem from one COX-2 molecule to another during crystallization can account for the observed turnover.

The orientation of PGH₂ in this structure agrees with the predicted orientation of product in the active site (Fig. 3b). The carboxylate end of PGH₂ is proximal to side chains of Arg 120 and Tyr 355, while its ω -end passes between Tyr 385 and Ser 530, entering the top channel of the active site. Arginine 120 is involved in AA binding in COX-1 (ref. 19), but its precise role in COX-2 remains unclear²⁰. The ω -end of the product makes a close contact with Gly 533, which, when mutated to either valine or leucine, completely abolishes cyclooxygenase activity when AA (20 carbon atoms) is the substrate²¹. Interestingly, the oxidation of shorter fatty-acid substrates, linolenic and stearidonic acids (both 18 carbon atoms) is unaffected with the Gly 533 variants. These results are consistent with the experimentally observed position of the ω -end of PGH₂ in the top channel of the cyclooxygenase active site.

The PGH₂ endoperoxide ring is positioned by van der Waals interactions with the side chains of the conserved hydrophobic residues Phe 381, Leu 384, Tyr 385 and Trp 387, the latter residue providing the bulk of the interactions. We constructed a Trp387Phe variant of COX-2 that has an altered product profile as compared with that of wild-type enzyme. The mutant enzyme produces 19-fold less prostaglandin than wild type (86 versus 5 pmol per pmol

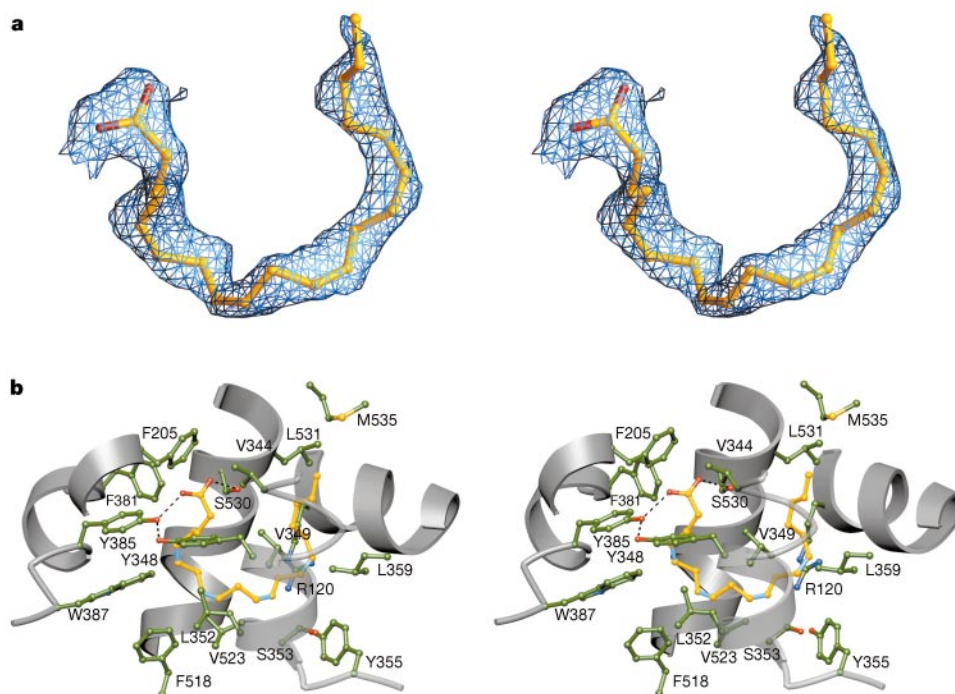


Figure 1 Arachidonic acid bound to the cyclooxygenase active site of the His207Ala variant of COX-2 at 2.4 Å resolution. **a**, Stereo image of the experimental $2(F_{\text{obs}}) - (F_{\text{calc}})$ difference electron density map (contoured at 1.0σ) before the inclusion of the substrate in

the phases. **b**, Stereo diagram of arachidonate bound to the cyclooxygenase active site. All protein side chains within van der Waals contact are shown; double bonds of AA are shown in blue.

enzyme), while production of 11-HETE, is essentially unchanged (5.36 versus 3.9 pmol 11-HETE per pmol enzyme). Consistent with the crystal structure, these data demonstrate the critical role played by Trp 387 in stabilization of the substrate conformation that leads to prostaglandin synthesis. Interestingly, the equivalent mutation in COX-1 resulted in a 50% reduction of cyclooxygenase activity²².

Most of the active site is unchanged by PGH₂ binding, but two side-chain rotamers are observed for Leu 384 and Tyr 385, one

consistent with AA binding and the other with PGH₂ binding. The side chain of Leu 384 rotates approximately 60° to accommodate the PGH₂ ring. In the absence of haem, the Tyr 385 side chain rotates to occupy the vacant haem binding pocket (Fig. 3b). This allows the PGH₂ to move closer to Tyr 385. Modelling studies suggest that displacement of PGH₂ towards Arg 120 by ~1.5 Å would allow restoration of the Tyr 385 side-chain position. Sufficient space in the active site exists to accommodate this translated

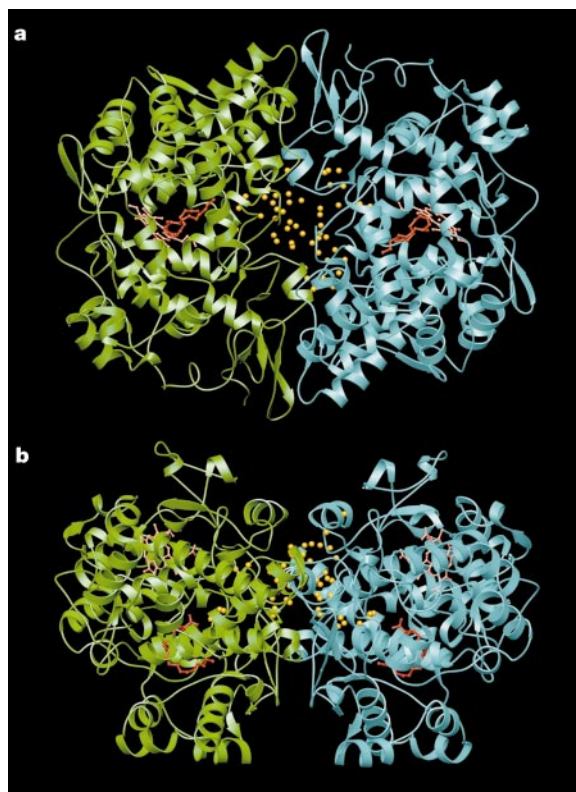


Figure 2 COX-2 dimer interface solvent channel. The two monomers (coloured green and blue) of the COX-2 dimer are shown from the membrane face (**a**) or side (**b**). Cyclooxygenase and peroxidase active sites are marked by AA (red) and haem molecules

(orange, superimposed onto the H207A-AA structure), respectively. Solvent molecules are shown as yellow spheres.

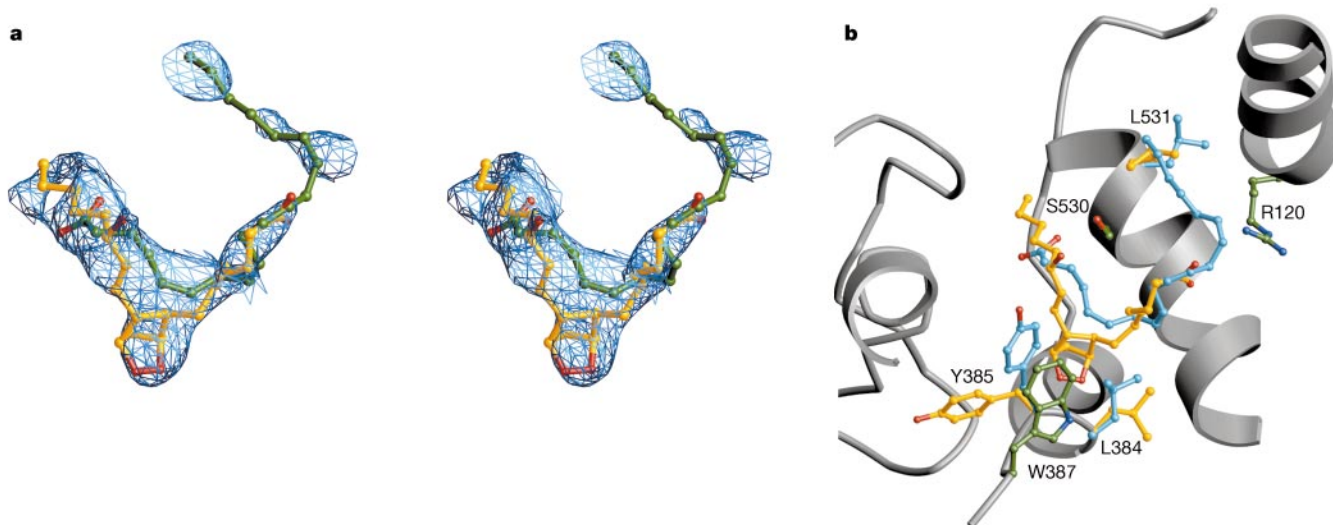


Figure 3 Arachidonate and prostaglandin bound to the cyclooxygenase active site of wild-type COX-2. **a**, Stereo image of the experimental ($(F_{\text{obs}}) - (F_{\text{calc}})$) difference electron density map (contoured at 1.7σ , additional contours in supplemental materials) before the

inclusion of AA (green trace) or PGH₂ (gold trace) in the phases. **b**, Diagram of PGH₂ (gold) and AA (cyan) bound to the cyclooxygenase active site. Mobile side chains are coloured gold and cyan to match their PGH₂ and AA bound positions, respectively.

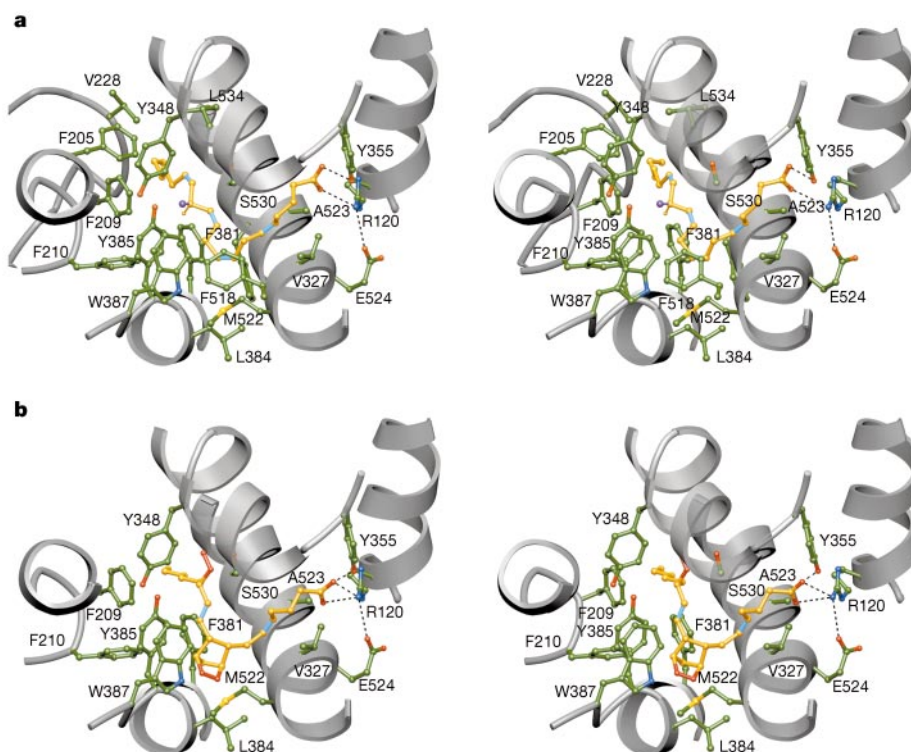


Figure 4 Stereo diagram of the models of AA (**a**) and PGH₂ (**b**) bound at the cyclooxygenase active site. Dashed lines indicate hydrogen bonds or ion pairs formed. The double bonds of the AA and PGH₂ are coloured blue. Protein side chains shown are

within van der Waals contact of the ligand. The C13 *pro(S)*-hydrogen (purple) and *pro(R)*-hydrogen (gold) of AA are shown for reference.

position of PGH₂ without movement of additional protein residues. This translated position of PGH₂ should represent its position in the haem-containing enzyme. Another consequence of the translation of the prostaglandin is that an ion pair forms between the side chain of Arg 120 and the carboxylate of PGH₂.

On the basis of the structure of PGH₂ bound at the cyclooxygenase active site, we propose a working model for the mechanistically competent conformation of AA that leads to prostaglandin synthesis (Fig. 4). Steps leading to the bicyclic ring formation can only be modelled on the basis of steric considerations and mutagenesis. In our model, the carboxylate of AA is coordinated by Arg 120 and Tyr 355. The C13 *pro(S)*-hydrogen atom of AA is positioned 2.5 Å from Tyr 385, allowing the abstraction of the hydrogen^{8,16}. Hydrogen abstraction generates a pentadienyl radical centred on C11 that is quenched by O₂ addition, forming a peroxy radical²³. Subsequent attack of oxygen at C11 is only possible from the side of the substrate that is opposite the side of hydrogen abstraction, because of steric blockage by protein residue Phe 381. These geometrical restrictions are consistent with the observed antarafacial stereochemistry of oxygen attack. The oxygen radical at C11 then cyclizes by attacking the C9 position.

The extremely compact environment of the bridged cyclopentane ring of PGG₂/PGH₂ and the shape of the pocket into which the ω-end of the molecule must pack both contribute to the selection of the *trans* configuration for the second cyclization reaction that links C8 to C12. A *cis* linkage of these rings would force a clash between the carboxylate end of PG and helix 17 of COX-2 (residues 520–535). The stereochemical specificity of the oxygenation at the C15 position of AA is also directed by the surrounding protein structure. Oxygen attack at C15 results in formation of the (*S*)-isomer, because the opposite side of the PGG₂ tail is blocked by residues Tyr 348, Phe 381, Tyr 385 and Ser 530.

The inherent biochemical properties of membrane proteins have presented daunting challenges to structural biology. The high-

resolution complex of AA bound to the enzyme reveals an unexpected binding orientation of arachidonic acid. This may be indicative of a non-productive binding mode for substrate in the cyclooxygenase active site. The structure of PGH₂ bound to COX-2 provides useful insights into the stereochemical course of the cyclooxygenase reaction. On the basis of this structure and published biochemical data, we can propose a working model for the binding of arachidonic acid to COX-2 that leads to prostaglandin formation. □

Methods

Crystallography

Recombinant murine COX-2 co-crystals for both experiments were grown as described²⁴ except that 1 mM AA was used in place of inhibitor. Data were integrated and scaled with DENZO and SCALEPACK²⁵. The His207Ala-AA co-crystal belongs to space group I222 ($a = 181.3$ Å, $b = 134.0$ Å, $c = 124.8$ Å) with one dimer per asymmetric unit. The 2.4 Å dataset (98.8% complete, $R_{\text{sym}} = 7.8\%$) was 3.6-fold redundant. The refinement statistics reflect excellent stereochemistry (r.m.s. deviation from ideal bond lengths, 0.010 Å) and agreement with the data ($R_{\text{free}} = 23.7\%$, $R = 20.6\%$). The AA-PGH₂ co-crystal belongs to space group P2₁2₁2 ($a = 180.2$ Å, $b = 134.8$ Å, $c = 122.5$ Å), containing two COX-2 dimers per asymmetric unit. The 3.0 Å dataset (96.2% complete, $R_{\text{sym}} = 11.2\%$) was 3.4-fold redundant. Both structures were solved by difference Fourier methods, starting with the published COX-2 structure⁷. We used standard building and refinement techniques^{26–29}. For the 3.0 Å resolution structure, refinement of AA and PGH₂ positions was performed independently, assuming unit occupancy. The resolution of the data does not permit occupancy refinement; however, we estimate that PGH₂ accounts for 30–50% of the population. The R_{free} and R -factor for the model are 32.1% and 26.7%, respectively. Liquid chromatography coupled mass spectrometry (LC/MS/MS) of the crystallization experiments showed that the samples contain large quantities of PGE₂ (the product of spontaneous conversion of PGH₂), but a positive control profile of PGH₂ could not be obtained from any commercial samples. PGH₂ is therefore inferred to be the prostaglandin species in the crystal because it is readily distinguished from PGE₂ crystallographically and is required for the formation of the latter. We co-crystallized COX-2 with AA, alternate substrates, products and analogues in the presence of haem; however, all such experiments either failed to yield crystals or resulted in crystals that were unsuitable for diffraction studies. Therefore, the current structure is the best surrogate of prostaglandin bound to the holoenzyme available.

Product profile

Assays were performed on purified wild-type and Trp387Phe variant murine COX-2 protein. Protein samples (750 nM) in 190 µl Tris-HCl buffer (100 mM, pH 8.0; 500 µM phenol at 37 °C) were treated with 100 µM [14 C]arachidonic acid (in 10 µl ethanol) for either 10 min or 1 h at 25 °C. Incubations were terminated by addition of 400 µl stop solution (80.6 µl ether, 11.4 µl methanol, 2.8 µl 1 M citric acid) after 1–3 h. The organic phase was removed and dried under nitrogen. Samples were then purged with argon and stored at –20 °C. To determine the products synthesized in the test samples, 5-, 11-, 12- and 15-HETE standard elution profiles were obtained. 1.6 µg of each standard was extracted and dried under nitrogen in the manner described above, resuspended in 500 µl methanol and placed in eppendorf tubes in 50 µl aliquots before purging with argon. HPLC conditions were modified from previous work³⁰; we used a 5-mm Beckman Ultrasphere ODS C₁₈ column (4.6 mm × 25 cm) run under reverse phase conditions at 1 ml min⁻¹ with a mixture of the following buffers: buffer A, 0.1% v/v glacial acetic acid in water; buffer B, 0.1% v/v glacial acetic acid in acetonitrile. Initial conditions used were 70% A/30% B for 5 min then the gradient was changed linearly to 40% A/60% B over 30 min. After a 10-min isocratic period, the gradient was changed to 25% A/75% B in a linear manner over 15 min. This was followed by a 10-min isocratic period before changing the gradient to 100% B linearly over 10 min followed by a 5-min isocratic period. Product quantification was determined by use of an online liquid scintillation counter (IN/US Systems, β-RAM 2B) and a Varian 2050 detector at 235 nm to detect product containing a conjugated double bond.

Received 4 January; accepted 29 February 2000.

- Marnett, L. J., Rowlinson, S. W., Goodwin, D. C., Kalgutkar, A. S. & Lanzo, C. A. Arachidonic acid oxygenation by COX-1 and COX-2: Mechanisms of catalysis and inhibition. *J. Biol. Chem.* **274**, 22903–22906 (1999).
- Smith, W. L., Garavito, R. M. & DeWitt, D. L. Prostaglandin endoperoxide H synthases (cyclooxygenases)-1 and -2. *J. Biol. Chem.* **271**, 33157–33160 (1996).
- Kulmacz, R. J. Cellular regulation of prostaglandin H synthase catalysis. *FEBS Lett.* **430**, 154–157 (1998).
- Herschman, H. R. Prostaglandin synthase 2. *Biochim. Biophys. Acta* **1299**, 125–140 (1996).
- Picot, D., Loll, P. J. & Garavito, R. M. The X-ray crystal structure of the membrane protein prostaglandin H₂ synthase-1. *Nature* **367**, 243–249 (1994).
- Luong, C. *et al.* Flexibility of the NSAID binding site in the structure of human cyclooxygenase-2. *Nature Struct. Biol.* **3**, 927–933 (1996).
- Kurumbail, R. G. *et al.* Structural basis for selective inhibition of cyclooxygenase-2 by anti-inflammatory agents. *Nature* **384**, 644–648 (1996); erratum *ibid* **385**, 555 (1997).
- Hamberg, M. & Samuelsson, B. On the mechanism of the biosynthesis of prostaglandins E-1 and F-1-α. *J. Biol. Chem.* **242**, 5336–5343 (1967).
- Dietz, R., Nastainczyk, W. & Ruf, H. H. Higher oxidation states of prostaglandin H synthase. Rapid electronic spectroscopy detected two spectral intermediates during the peroxidase reaction with prostaglandin G₂. *Eur. J. Biochem.* **171**, 321–328 (1988).
- Smith, W. L., Eling, T. E., Kulmacz, R. J., Marnett, L. J. & Tsai, A. Tyrosyl radicals and their role in hydroperoxide-dependent activation and inactivation of prostaglandin endoperoxide synthase. *Biochemistry* **31**, 3–7 (1992).
- Kulmacz, R. J., Pendleton, R. B. & Lands, W. E. Interaction between peroxidase and cyclooxygenase activities in prostaglandin-endoperoxide synthase. Interpretation of reaction kinetics. *J. Biol. Chem.* **269**, 5527–5536 (1994).
- Landino, L. M., Crews, B. C., Gierse, J. K., Hauser, S. D. & Marnett, L. J. Mutational analysis of the role of the distal histidine and glutamine residues of prostaglandin-endoperoxide synthase-2 in peroxidase catalysis, hydroperoxide reduction, and cyclooxygenase activation. *J. Biol. Chem.* **272**, 21565–21574 (1997).
- Word, J. M. *et al.* Visualizing and quantitating molecular goodness-of-fit: small-probe contact dots with explicit hydrogens. *J. Mol. Biol.* **285**, 1711–1733 (1998).
- Karthein, R., Dietz, R., Nastainczyk, W. & Ruf, H. H. Higher oxidation states of prostaglandin H synthase. EPR study of a transient tyrosyl radical in the enzyme during the peroxidase reaction. *Eur. J. Biochem.* **171**, 313–320 (1988).
- Shimokawa, T., Kulmacz, R. J., DeWitt, D. L. & Smith, W. L. Tyrosine 385 of prostaglandin endoperoxide synthase is required for cyclooxygenase catalysis. *J. Biol. Chem.* **265**, 20073–20076 (1990).
- Hamberg, M. Stereochemistry of oxygenation of linoleic acid catalyzed by prostaglandin-endoperoxide H synthase-2. *Arch. Biochem. Biophys.* **349**, 376–380 (1998).
- Eling, T. E., Glasgow, W. C., Curtis, J. F., Hubbard, W. C. & Handler, J. A. Studies on the reduction of endogenously generated prostaglandin G₂ by prostaglandin H synthase. *J. Biol. Chem.* **266**, 12348–12355 (1991).
- Ueno, R., Shimizu, T., Kondo, K. & Hayaishi, O. Activation mechanism of prostaglandin endoperoxide synthetase by hemoproteins. *J. Biol. Chem.* **257**, 5584–5588 (1982).
- Bhattacharyya, D. K., Lecomte, M., Rieke, C. J., Garavito, M. & Smith, W. L. Involvement of arginine

- 120, glutamate 524, and tyrosine 355 in the binding of arachidonate and 2-phenylpropionic acid inhibitors to the cyclooxygenase active site of ovine prostaglandin endoperoxide H synthase-1. *J. Biol. Chem.* **271**, 2179–2184 (1996).
- Rieke, C. J., Mulchak, A. M., Garavito, R. M. & Smith, W. L. The role of arginine 120 of human prostaglandin endoperoxide H synthase-2 in the interaction with fatty acid substrates and inhibitors. *J. Biol. Chem.* **274**, 17109–17114 (1999).
- Rowlinson, S. W., Crews, B. C., Lanzo, C. A. & Marnett, L. J. The binding of arachidonic acid in the cyclooxygenase active site of mouse prostaglandin endoperoxide synthase-2 (COX-2): a putative L-shaped binding conformation utilizing the top channel region. *J. Biol. Chem.* **274**, 23305–23310 (1999).
- Hsi, L. C. *et al.* Trp387 and the putative leucine zippers of PGH synthases-1 and -2. *J. Lipid. Mediat.* **6**, 131–138 (1993).
- Tsai, A., Palmer, G., Xiao, G., Swinney, D. C. & Kulmacz, R. J. Structural characterization of arachidonyl radicals formed by prostaglandin H synthase-2 and prostaglandin H synthase-1 reconstituted with mangano protoporphyrin IX. *J. Biol. Chem.* **273**, 3888–3894 (1998).
- Stevens, A. M. *et al.* Crystallization of recombinant cyclo-oxygenase-2. *J. Cryst. Growth* **196**, 350–355 (1999).
- Otwinski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Brünger, A. T. X-PLOR version 3.1: A system for X-ray crystallography and NMR. Yale University Press, New Haven, CT, USA (1992).
- Jiang, J. -S. & Brünger, A. T. Protein hydration observed by X-ray diffraction. Solvation properties of penicillopepsin and neuraminidase crystal structures. *J. Mol. Biol.* **243**, 100–115 (1994).
- Read, R. J. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A* **42**, 140–149 (1986).
- Jones, T. A., Zou, J. -Y. & Cowan, S. W. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Capdevila, J. H. *et al.* The catalytic outcomes of the constitutive and the mitogen inducible isoforms of prostaglandin H₂ synthase are markedly affected by glutathione and glutathione peroxidase(s). *Biochemistry* **34**, 3325–3337 (1995).

Acknowledgements

We are grateful to M. C. Walker, R. Fletterick, T. Rydel, H.-S. Shieh, J. Pierce, P. C. Isakson and K. Seibert for substantive discussions. LC-MS experiments were performed by J. Muhammad. Figures were generated with RIBBONS (M. Carson). S.W.R. is the recipient of an Australian National Health and Medical Research Council C.J. Martin Fellowship.

Correspondence and requests for materials should be addressed to R.G.K. (e-mail: ravi.g.kurumbail@monsanto.com). Coordinates for structures and models have been deposited in the Protein Data Bank under accession codes 1CVU, 1DCX, 1DD0 and 1DDX.

erratum

The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes

Yin-Long Qiu, Jungho Lee, Fabiana Bernasconi-Quadroni, Douglas E. Soltis, Pamela S. Soltis, Michael Zanis, Elizabeth A. Zimmer, Zhiduan Chen, Vincent Savolainen & Mark W. Chase

Nature **402**, 404–407 (1999)

The received and accepted dates of this Letter were inadvertently omitted. They should have read: Received 6 September; accepted 13 October 1999.

Why not take one for a spin?

The latest in centrifuges — loads of *g*, masses of speed, but not much noise.

Genofuge

Techne

www.techne.co.uk

Hush-hush high-speed centrifugation

The Genofuge microcentrifuge is powered by a maintenance-free brushless motor, claimed to give extremely quiet running, and to generate a minimal amount of heat. The unit features a ducting system for a continuous flow of ambient air to keep the rotor and sample cool. For below ambient conditions Genofuge can be used in refrigerated environments. Speed varies from 1,000 to 14,000 r.p.m. to suit microfiltration, separation of proteins and nucleic acids, and other applications.

Reader Service No. 100

K Prima-15R

Composite Rotor Inc.

www.composite.com

A new angle on carbon fibre

The KOMPspin Prima-15R refrigerated tabletop centrifuge is built for large capacity and quiet running. The unit spins 1.5 ml, 50 ml, and 85 ml tubes at up to 15,000 r.p.m. (27,000g) with its non-corroding carbon fibre angle rotor, 1–100 ml and 15/50 ml conical tubes in swinging buckets, a four-place 96-well microtitre plates or six-place slides in the cyto rotor. Power is by brushless induction drive.

Reader Service No. 101

Clear-Spin

Seward

www.seward.co.uk

Here's one to look into

The Clear-Spin microcentrifuge is intended to simplify the manipulation of hazardous samples. The rotor disc doubles as a tube rack, so sample preparation can be performed while the tubes remain in the rotor. The rotor case is hermetically sealed and made from transpar-

ent nylon so that any spillage can be spotted before lab personnel are exposed to risk.

Reader Service No. 102

Centrifuge 5417 C

Eppendorf

www.eppendorf.com

Running hot and cold...

The Eppendorf 5417 C, and the refrigerated 5417 R, now have an embossed foil keypad and an extended range of rotors. Microcentrifuge tubes spin-up to 25,000g almost instantly in a fixed-angle rotor, and a PCR rotor accommodates eight PCR tube strips. Other rotors supplied are the familiar 30-position fixed-angle rotor for 1.5/2.0 ml tubes (also available aerosol-tight) and a swing-bucket rotor.

Reader Service No. 103

Centra CL3

IEC

www.labcentrifuge.com

Wait for it...

CL3 Series centrifuges are available in refrigerated and non-refrigerated versions. Recent refinements include a 99-protocol storage facility, speed setting to the nearest 10 r.p.m. and automatic calculation of *g* force. The "at speed timer" assures that the timed run commences only when the rotor reaches the required *g* force. The CL3 accommodates a variety of tubes including 15 ml, 50 ml, 225 ml conical tubes, microtubes, microplates microscope slides and filtration devices.

Reader Service No. 104

KR4.22 centrifuge

Jouan

www.jouaninc.com

A calculating centrifuge

The Jouan KR 4.22 combines high capacity and high centrifugal force. It can process large numbers of microtitration plates and columns

at above 5,000g. The proven high torque direct-drive motor is unchanged in the latest models. A digital control system allows the user to control centrifugation precisely. First the centrifuge calculates the total separating force (*g*-time) applied to samples in a model run, then the run can be regulated to this value. The results can be transferred electronically to Jouan's data acquisition system.

Reader Service No. 105

TJ-25 benchtop

Beckman Coulter, Inc.

www.beckman.com

A 'personal' centrifuge for general lab use

Standard equipment on the TJ-25 high-speed refrigerated benchtop centrifuge includes a high-*g*-force (5,520g), high-volume, swinging-bucket TS-5.1-500 rotor, taking four 500-ml samples. This new rotor accommodates adapters for a range of tubes, including 50 ml and 15 ml conical tubes. An optional microtitre plate system for this rotor allows the use of up to four deep-well plates or eight standard plates. An optional biosafety lid for the rotor facilitates biocontainment.

Reader Service No. 106

Heraeus Multifuge

Kendro Laboratory Products

www.kendro.com

Think big

Designed for high sample throughput, the Heraeus Multifuge has a tube capacity that the makers claim is the largest in its class. A swinging bucket rotor and a new windshield rotor offer tube capacity up to 4 × 1,000 ml and 7,000g. A wide range of adapters offers compatibility with most common tube types.

Reader Service No. 107

These notes are compiled in the Nature office from information provided by the manufacturers.



Seward Clear-Spin: a touch of glass, well, nylon.

ADVERTISEMENTS

Custom Peptides & Polyclonal Antibodies

Please contact us about complete services for your custom peptide and polyclonal antibody needs.

Research Genetics
U.S. or Canada 800-533-4363
FAX 256-536-9016

Check out our homepage at
<http://www.resgen.com>

Look no further

for the best value in

Oligo Synthesis

Competitive pricing
Routine 24 hour dispatch • Every Oligo QC verified by PAGE
Rigorous batch QC by PCR and Mass Spec
Web ordering and tracking • Local Freephone numbers
A great deal in any language!!

SIGMA GENOSYS (+44) (0) 1223 839000 • fax (+44) (0) 1223 838200
email: info@sigma-genosys.co.uk • www.sigma-genosys.co.uk